

CDM-EB67-A06-GUID

Guideline

Sampling and surveys for CDM project activities and programmes of activities

Version 04.0



United Nations
Framework Convention on
Climate Change

TABLE OF CONTENTS	Page
1. INTRODUCTION	4
2. SCOPE, APPLICABILITY, AND ENTRY INTO FORCE	4
2.1. Scope and applicability	4
2.2. Entry into force.....	4
3. NORMATIVE REFERENCES.....	4
4. DEFINITIONS	4
5. COMMON TYPES OF SAMPLING APPROACHES	5
5.1. Simple random sampling	5
5.2. Stratified random sampling	5
5.3. Systematic sampling.....	6
5.4. Cluster sampling.....	6
5.5. Multi-stage sampling.....	7
6. RECOMMENDED OUTLINE FOR A SAMPLING PLAN.....	11
6.1. Sampling design	11
6.2. Data to be collected	11
6.3. Implementation	12
7. RECOMMENDED PRACTICES FOR UNBIASED ESTIMATES OF SAMPLED PARAMETERS.....	12
8. RECOMMENDED EVALUATION CRITERIA FOR DOE VALIDATION.....	14
9. RECOMMENDED PRACTICES FOR CONDUCTING SURVEY AND DATA COLLECTION.....	15
9.1. Survey and data collection methods by PPs.....	15
9.1.1. Physical on-site visit (face-to-face interview)	19
9.1.2. Remote survey.....	19
9.2. Survey and data collection methods in relation to validation/verification purposes by DOEs.....	21
9.2.1. Recommended evaluation criteria for DOE validation of survey and data collection methods used by PPs.....	21
9.2.2. Survey and data collection methods for use by DOEs for validation/verification of sample records of PPs.....	23

9.3.	Selecting a survey and data collection method and preference for use	23
9.4.	Data processing and report generation	24
9.4.1.	Data processing	24
9.4.2.	Report generation	25
APPENDIX 1.	BEST-PRACTICE EXAMPLES FOR SAMPLE SIZE CALCULATIONS	26
APPENDIX 2.	BEST-PRACTICE EXAMPLES FOR A SINGLE SAMPLING PLAN FOR A HOMOGENEOUS POA.....	65
APPENDIX 3.	BEST-PRACTICE EXAMPLES FOR A SINGLE SAMPLING PLAN FOR A HETEROGENEOUS POA USING A STRATIFIED SAMPLING APPROACH	75
APPENDIX 4.	BEST-PRACTICE EXAMPLES FOR RELIABILITY CALCULATIONS	93
APPENDIX 5.	BEST-PRACTICE EXAMPLES – ACCEPTANCE SAMPLING	108
APPENDIX 6.	SAMPLE SIZE CALCULATOR.....	113

1. Introduction¹

1. This document includes generic approaches for sampling and surveys applied to clean development mechanism (CDM) projects and programmes of activities (PoAs).

2. Scope, applicability, and entry into force

2.1. Scope and applicability

1. This document describes common types of sampling approaches and includes a recommended outline for a sampling plan, recommended practices for unbiased estimates of sampled parameters and recommended evaluation criteria for designated operational entity (DOE) validation besides several best-practice examples covering large and small-scale project activities and PoAs. It also provides examples for checking the reliability of data collected through sample surveys.
2. Furthermore, it covers the following items:
 - (a) Methods, if any, to deal with missed reliability targets without compromising conservative estimates for emission reduction;
 - (b) Best-practice examples for DOE validation/verification for sampling and surveys.

2.2. Entry into force

3. The date of entry into force of the revision is the date of the publication of the EB 86 meeting report on 16 October 2015.

3. Normative references

4. The following referenced document is indispensable for the application of these guidelines:
 - (a) "Standard for sampling and surveys for CDM project activities and programme of activities" (hereinafter referred to as the Sampling standard);

4. Definitions

5. The definitions contained in the Glossary of CDM terms and the Sampling standard shall apply.

¹ For this document, information on, inter alia, the background, purpose, key issues and solutions, and recommendations to the Board, is contained in the cover note of annex 11 "Revision of regulatory documents due to introduction of or changes to provisions related to PoAs" (CDM-EB75-AA-A11), which is also included on the annotated agenda of the seventy-fifth meeting of the Board.

5. Common types of sampling approaches

6. This section provides a summary² of some of the most common types of sampling approaches and typical situations where each is recommended. Formulas for calculating standard errors of estimates from each sampling technique, confidence intervals and associated sample sizes are provided in the reference texts cited at the end of this report. The provided sampling information primarily relates to determining point estimates of average (mean) values of a parameter.

5.1. Simple random sampling

7. A *simple random sample* is a subset of a population (e.g. villages, individuals, buildings, pieces of equipment) chosen randomly, such that each element (or unit) of the population has the same probability of being selected. The sample-based estimate (mean or proportion) is an unbiased estimate of the population parameter.
8. Simple random sampling is conceptually straightforward and easy to implement – provided that a sampling frame of all elements of the population exists. Its simplicity makes it relatively easy to analyse the collected data. It is also appropriate when only minimum information of the population is known in advance of the data collection.
9. Simple random sampling is suited to populations that are homogeneous. In many instances a large population size and dispersed nature of population may cause a lack of homogeneity, while in some cases those factors may have relatively low impact on homogeneity (e.g. a large number of biogas digesters located in varying altitudes and temperature zones may be less conducive for simple random sampling to determine the average amount of biogas production per digester, while the usage hours of light bulbs across wide geographic areas and among large populations with similar socioeconomic circumstances connected to a single or similar grid/s may be sufficiently homogeneous for simple random sampling). The costs of data collection under simple random sampling could be higher than other sampling approaches when the population is large and geographically dispersed.

5.2. Stratified random sampling

10. When the population under study is not homogeneous but instead consists of several sub-populations which are known (or thought) to vary, then it is better to take a simple random sample from each of these sub-populations separately. This is called *stratified random sampling*. The sub-populations are called the strata. When considering stratified random sampling it is important to note that when identifying the strata no population element can be excluded and every element must be assigned to only one stratum. For example, the population of participants in a commercial lighting programme might be grouped according to building type (e.g. restaurants, food stores, and offices).
11. Stratified random sampling is most applicable to situations where there are obvious groupings of population elements whose characteristics are more similar within groups than across groups (e.g. restaurants are likely to be more similar to one another in terms of lighting use than they are to offices or food stores). It requires that

² See table 1 for advantages and disadvantages of each sampling approach.

the grouping variable be known for all elements in the sampling 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.

12. Stratification helps to ensure that estimates of a population characteristic are accurate, especially if there are differences amongst the strata. For example, if lighting use within office buildings tends to be lower (on average) than in food stores then this can be taken into account when estimating the overall average number of hours of operation. Equally, if the cases within each stratum are more homogeneous than across strata, then the estimated number of hours of operation will be more precise than if a simple random sample of the same size had been taken.

5.3. 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:

$$k = \text{population size } (N) / \text{sample size } (n)$$

14. Using this procedure, each element in the population has a known and equal probability of selection. The project participant shall 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.
15. Systematic sampling is applicable in a number of situations. If there is a natural ordering or flow of subjects 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 as they are produced. 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.

5.4. Cluster sampling

16. *Clustered sampling* refers to a technique where the population is divided into sub-groups (clusters), and the sub-groups are randomly selected (sampled), rather than the individual elements which are to be studied. The data are then collected on all the individual elements in the selected sub-groups.
17. Cluster sampling is used when “hierarchical” groupings are evident in a population, such as villages and households within villages, or buildings and appliances within buildings. For example, suppose a project installs high-efficiency motors in new apartment buildings, with several motors typically in each building. In order to estimate 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.
18. 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. apartment buildings), and there is no expectation that the motors in each building are more homogeneous than the overall population of efficient motors.

19. Cluster sampling is useful when there is no sampling frame at the lowest level of the hierarchy but there is one at the cluster level, as in the case above where a ready list of all motors would not be available, but a list of all new apartment buildings would be.
20. 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.
21. 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 meter more pieces of equipment (sample more clusters) to achieve the same level of precision as the simple random sampling, but the reduction in cost and other benefits may more than offset this apparent increase in effort.

5.5. Multi-stage sampling

22. *Multi-stage sampling* is a more complex form of cluster sampling. Measuring all the elements in the selected clusters may be prohibitively expensive, or not even necessary. In multi-stage sampling, the cluster units are often referred to as primary sampling units and the elements within the clusters secondary sampling units. In contrast to cluster sampling where all of the secondary units are measured, in multi-stage sampling data are collected for only a sample of the secondary units.
23. For example, in a study of efficient lighting, if the operation hours of motors within any one building are thought likely to be similar across all motors then – especially if the cost of measuring them is relatively high – there is not much to be gained by metering all of them. It might be better to draw a sample of buildings, and then only measure a sample of motors from within each selected building. 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.
24. 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.
25. So far, most of the methods above have been based on simple random sampling. Another option is to sample with probability proportional to size, and this is sometimes used in cluster sampling where clusters are of different sizes, or in multi-stage sampling.
26. There are therefore many variations in methods in applying multi-stage sampling. If the number of secondary units in each primary unit is not known in the sampling 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. 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,

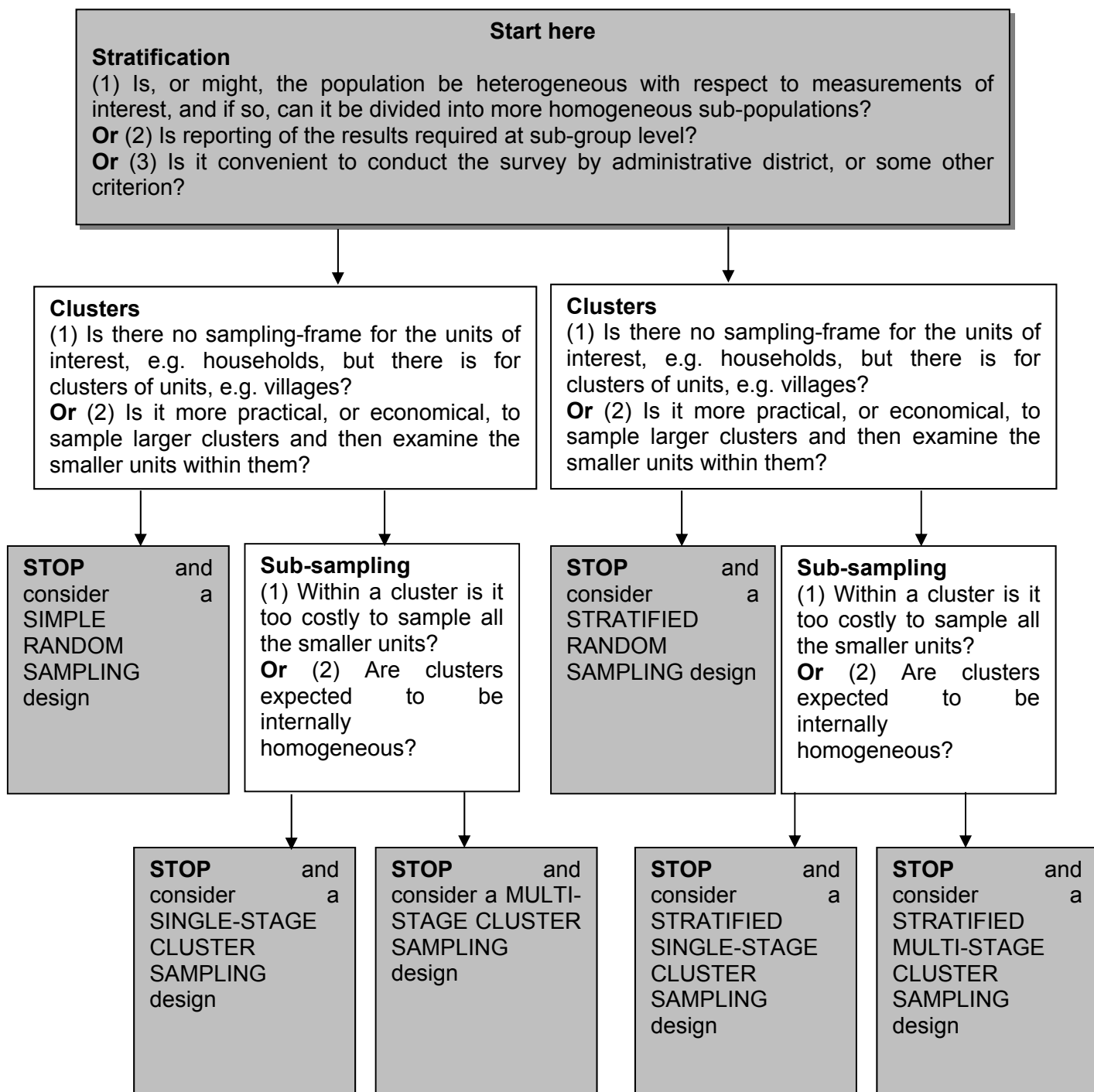
and the availability of information on the primary and secondary units in the sample frame.

27. Table 1 below indicates advantages and disadvantages of various sampling schemes in a summary form and figure 1 below shows a flow chart for selecting a survey design.

Table 1 Advantages and disadvantages of different sampling schemes

Sampling Scheme	Advantages	Disadvantages
<p>Simple Random Sampling: Taking a random sample from the whole population.</p>	<p>Easiest method to understand and therefore use. Suitable if there is little heterogeneity amongst the units being sampled</p>	<p>Requires knowledge of 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. Only suitable if the population being studied is relatively homogeneous with respect to the parameter being studied</p>
<p>Systematic Sampling: Taking a sample every n units</p>	<p>Easy to apply. Commonly used as it ensures there is always sufficient distance between samples</p>	<p>Leads to units being spread out over a large geographic area. Such a geographic distribution can often be costly</p>
<p>Stratified Random Sampling: Randomly sampling a different number of units from each strata according to the weight of each strata in the population</p>	<p>Improves the precision of the estimate (compared to simple random sampling) if there are differences between the strata</p>	<p>Complicated to calculate. What the stratification factors should be is not always obvious</p>
<p>Cluster Sampling: Sampling every unit in a sample of n clusters from the population</p>	<p>The most economical form of sampling as units are all grouped according to one criterion (often geographical). Sometimes the only approach, since a list of all households may not be available, only a list of villages. Once the villages have been selected, the households can be sampled. It saves time at a management level</p>	<p>Results are not normally so 'good' (i.e. standard errors of estimates tend to be high due to homogeneity of characteristics in the subgroup sampled). But a larger sample can help to compensate for this</p>
<p>Multi-stage Sampling: Randomly sampling a number of units within a number of randomly selected clusters</p>	<p>Enables sampling approach at two levels. Can compare different scenarios – number of clusters and number of units within the clusters – in order to find most cost-efficient and reliable scenario</p>	<p>Analysis and the sample size calculation are more difficult</p>

Figure 1. Flow chart for selecting a survey design



28. This flow chart helps the user to select one of six survey designs by answering yes or no to a series of questions. Readers are advised to consult section “9.1 Reasons for cluster sampling” under chapter 9 of Sampling Techniques by Cochran (1977)³ to better understand the connotations with which terms such as practical, economical and costly are used in this flow chart. Cochran also provides further details on all of the sampling techniques mentioned in this flow chart.

6. Recommended outline for a sampling plan

29. The sampling plan should contain information relating to: (A) sampling design; (B) data to be collected; and (C) implementation plan.

6.1. Sampling design

30. *Objectives and reliability requirements* describes the objective of the sampling effort, the timeframe, and the estimated parameter value(s). Identify the sampling requirements (applicable CDM methodology or sampling standards) and the confidence/precision criteria to be met. For example, the objective is determining the mean monthly value of parameter “X” during the crediting period, and with a 90/10 confidence/precision.
31. *Target population* defines the target population, and describes any particular features associated with it.
32. *Sampling method* selects and describes the sampling method, e.g. simple random sampling, stratified sampling, cluster sampling. Strata or clusters shall be clearly identified if sampling other than simple random sampling is to be used.
33. *Sample size* addresses and justifies the estimated target number of “units” – pieces of equipment, solar cookers, buildings, motors, log-books, etc. – which are to be studied (i.e. the sample size). The justification shall include the parameter of interest, the value it is expected to take and an estimate of the variance associated with the data, as well as the level of confidence and precision (note that if the parameter of interest is a proportion, or a percentage, then there is no need to specify a variance estimate).
34. *Sampling frame* identifies or describes the sampling frame to be used. This shall agree with the information about the Target Population and Sampling Design above. For instance, if cluster sampling is to be used in a study of equipment in buildings, then the frame should be a listing of the buildings from which the sample will be selected.

6.2. Data to be collected

35. *Field measurements* identify all the variables to be measured and determine appropriate timing and frequency of the measurements. When the measurements are conducted only during limited time periods and are to be scaled up to the whole year, demonstrate that the parameter of interest is not subject to seasonal fluctuations or the time period selected is conservative or the necessary corrections are applied. Methods of measurement shall be described as appropriate;

³ Cochran, W.G. (1977). Sampling Techniques. Third edition, Wiley: New York.

36. *Quality assurance/Quality control* describe how to achieve good quality data, for example describe the procedures for conducting the data collection and/or field measurements 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. An overall quality control and assurance strategy shall be documented in the plan. This shall include a procedure for defining outliers and under what circumstances outlier data/measurements may be excluded and/or replaced.
37. *Analysis* describes how the data will be used.

6.3. Implementation

38. *Implementation plan* defines the schedule for implementing the sampling effort and identify the skills and resources⁴ required for data collection and the analyses.

7. Recommended practices for unbiased estimates of sampled parameters

39. Practitioners are expected to observe sound practices in designing samples and administering surveys and field measurements.⁵ Those practices include:
- (a) **Defining precisely the sampling objectives and target population and the measurements to be taken and/or data collected.** The sampling objectives will, for the most part, be concerned with estimation, i.e. estimating a characteristic (e.g. mean or percentage) of a population. Occasionally, the objective will be one of comparison, for example to compare the uptake in rural areas with that of urban areas:
 - (i) The target population is the “greater entity” to which the results from the survey sample are to be generalized, for example all new light fittings that are installed in new buildings in country X;
 - (ii) The information that will be collected will depend on the objectives, for example if a project needs to estimate the average number of hours of operation of a new efficient motor, then the data to be collected on each sampling unit is its number of hours of operation. Other measurements to be taken may relate to the characteristics of the strata, or clusters, or any other variable that may be relevant to the project objectives.
 - (b) **Deciding on the sampling design and the size of the sample.** This decision is based on the information provided above;
 - (c) **Developing the sampling frame.** A *sampling frame* is a complete listing of all individual units (elements, members) that can be considered as a

⁴ A general description of qualifications and experience of personnel who will be engaged should be provided, not necessarily listing specific names, qualification and experience.

⁵ 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.

representation of the whole population, and which can be used as a basis for selecting a sample, such as a list of all households in an area that have had solar cookers installed.⁶ In the case of cluster sampling or multi-stage sampling, the sampling frame is a complete listing of sub-groups of the study area/population⁷ which constitutes all the clusters or primary sampling units.

Without such a frame, or its equivalent, methods of sampling with assured properties such as unbiasedness are not available. The implementer of the survey effort shall compile a clear description of the target population, including those characteristics of the population which define membership. From the description and characteristic the implementer can then select a sampling frame;

- (d) **Randomizing cases and drawing sample.** The implementer should ensure that the sample is drawn at random from the sampling frame. This can be done using random number tables or using the random number generator of appropriate software. 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;
- (e) **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. The methods of survey and data collection and criteria to be used for selecting an appropriate survey and data collection method by the project implementer are described in section 9;
- (f) **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 intentional errors or unintentional 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;
- (g) **Minimizing non-response and adjusting for its effects.** The project implementer is expected to make all reasonable efforts to minimize non-response, to analyze potential bias arising from non-response, and to correct for any detected biases or losses in precision due to non-response. Field data collection protocols should specify procedures for multiple contacts to minimize non-response, require documentation of reasons for non-response, and prescribe corrective measures to compensate for its occurrence. Corrective measures may include over-sampling, replacing non-respondents with similar subjects, applying “correction factors” and imputing responses.

⁶ Such a listing shall be available for check during validation/verification but not necessarily included in the PDD documents.

⁷ 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.

8. Recommended evaluation criteria for DOE validation

40. The following questions and evaluation criteria serve as examples and should be utilized by DOEs to validate the proposed sampling plans:
- (a) Does the sampling plan present a reasonable approach for obtaining unbiased, reliable estimates of the variables?
 - (i) In terms of assessing reliability, are the elements of Objectives and Reliability Requirements complete? Do the requirements specified agree with those stated in the appropriate standards? If not, is there a reason why they are not met?
 - (ii) From all the different elements of the Design, is there any reason to suspect that the results from the activity will be biased? For instance, is the population under consideration only urban households? What about rural households? Might this cause a bias when the data are extrapolated to emission reductions?
 - (b) Is the population clearly defined, and how well does the proposed approach to developing the sampling frame represent that population?
 - (i) The population should be clear from the Target Population description. Whether or not the sampling frame is possible or appropriate will depend on the detail and the particular situation, for example if a map is going to be used, a question would be whether a map already exists, and how reliable it is. If a map does not exist, then who is going to create it?
 - (c) Is the proposed sampling approach clear?
 - (i) Is it clear which sampling method is being proposed? For example, is it simple random sampling, or some other method of sampling?
 - (ii) Does the method agree with the description of the population? Are there clusters or strata, and if so does it state what they are? For example, are they buildings, villages, etc.?
 - (d) 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?
 - (i) All of the information set out in the sampling plans should help answer this question. If not all information is provided then the question cannot be answered;
 - (ii) Is the target value for the population parameter reasonably anticipated?
 - (iii) Does the estimate of variability seem reasonable?
 - (e) Is the sample representative?
 - (i) Is it clear how the sample is to be selected? For example, is it to be selected randomly?

- (ii) Does the Plan indicate that the sampling frame will be kept (e.g. in hard copy or a computer file of screen shot copy), and that random numbers will be generated and these random numbers will then be used to select the sample?
- (f) Is the data collection/measurement method likely to provide reliable data given the nature of the parameters of interest and project, or is it subject to measurement errors?
 - (i) Are the methods of data collection clear and unambiguous? Are there questions which could be subject to respondent error due to sensitivity (e.g. “How much money do you spend on heating?”), lack of recall (e.g. “How many times did you buy fuel last year?”), and the like?
 - (ii) Are there questions that could be subject to measurement error? For example, is a particular measurement method known to under-record key data, such as the weight of bricks?
- (g) Are the procedures for the data measurements well defined and do they adequately provide for minimizing non-sampling errors?
 - (i) Is the quality control and assurance strategy adequate?
 - (ii) Are there mechanisms⁸ for avoiding bias in the answer?
- (h) Does the frame contain the information necessary to implement the sampling approach?
 - (i) Are the proposed skill sets, qualifications and experience of the personnel to be engaged to conduct sampling adequate?

9. Recommended practices for conducting survey and data collection

41. The following subsections contains guidance with regard to:

- (a) Methods of survey and data collection and criteria to be used for selecting an appropriate survey and data collection method by the PPs;
- (b) Methods of survey and data collection for validation/verification purposes and criteria to be used for selecting an appropriate survey method by the DOEs;
- (c) Data processing and report generation: minimum requirements for documentation.

9.1. Survey and data collection methods by PPs

42. A survey can be conducted in two ways: physical on-site visit (face-to-face) or remote survey. For data collection, different methodologies may be employed to collect data, for example, through hard-copy questionnaires, smartphones or tablet app modules

⁸ Mechanisms for avoiding non-sampling errors (bias) include good questionnaire design, well-tested questionnaires, possibly pilot testing the data collection.

connected to data clouds, data sensors, e-mail or web-based platform or SMS, telephone, mailing (post), etc.

43. The different methods available for survey and data collection are summarized in table 2.

Table 2 Survey and data collection methods and preference for use

Survey	Data collection method	Order of Preference	Comments	Anticipated response rate
Physical on-site visit (face-to-face interview)	Hard-copy questionnaires; or smartphone or tablet app modules	1st	<ul style="list-style-type: none"> • Higher quality data • Physical inspection by trained personnel • Longer more complex questions possible • Possible to collect more data • However, in the case of simple parameters such as survival rates (retention rates), well designed remote methods such as data loggers/sensors or pay-as-you-go systems may be equally effective or even perform better than physical on-site visits especially when the frequency of surveys is high 	High
Remote survey	Data sensor	1st	<ul style="list-style-type: none"> • Accurate data • May be used only under certain circumstances • Usually expensive unless it is part of the project design such as pay-as-you-go systems 	High
	Telephone interview	2nd	<ul style="list-style-type: none"> • Telephone necessary which may present sampling problems • Reduced response rate • Not suitable for long or complex questions • Identity of the respondent may not be clear 	Medium
	E-mail or web-based platform or SMS	3rd	<ul style="list-style-type: none"> • Access to a computer or other portable device required which may present sampling problems • Response rate may be very low – e-mails ignored, or filtered as spam • Different households may have different IT systems making the display of the questionnaire awkward • Questionnaire fatigue may require fewer well-structured questions, which need to be simple to answer • Further probing, or clarification of answer(s) not possible • Difficult to ascertain who completed the questionnaire 	Low

Survey	Data collection method	Order of Preference	Comments	Anticipated response rate
	Mailing (post) (questionnaires sent by regular mail)	3rd	<ul style="list-style-type: none"> • Respondent must be literate and capable of understanding and completing the questionnaire • Questionnaire fatigue may require fewer well-structured questions, which need to be simple to answer • Further probing, or clarification of answer(s) not possible • Difficult to ascertain who completed the questionnaire • Response rates are typically low 	Low

44. For the planning and selection of a survey and data collection method, the requirements of the CDM methodology shall be followed. In addition, a number of technical factors need to be considered, such as the target population, the frequency of the survey, type of parameter, the statistical design of the survey, response rates, and sources of measurement error. Also, practical aspects such as resources and costs to conduct a survey and data collection should be taken into account.

9.1.1. Physical on-site visit (face-to-face interview)

45. For the physical on-site visit (face-to-face interview), data is collected through (i) hard-copy questionnaires or (ii) smartphone or tablet app modules connected to data clouds. The physical on-site visit will allow the purpose of the study to be explained to the interviewed households, and therefore would be expected to reduce drop-out rates. Physical observation and recording is a clear advantage for on-site visits.
46. More efficient data collection devices such as smartphone or tablet questionnaires instead of hard-copy questionnaires (e.g. using app modules for surveys while visiting the households and filling in the digital questionnaires instead of filling in hard-copy questionnaires) are increasingly being used in surveys. Such computer-assisted personal interviewing (CAPI) has clear advantages over paper-based questionnaires when a large number of samples are involved, although for a small number of samples (e.g. under 15) it may be expensive and hence not practical to implement. For example, CAPI allows data validation checks to be enforced at the point of collection, so minimizing room for measurement errors.
47. Advanced data collection systems employ multiple mobile digital devices (smartphones and tablets, for example) that, soon after collection, load the captured data to a (cloud-based) data server for central storage that may also include data sorting, management and even processing or reporting. This way, storage is not relying on a single hand-held device that may be damaged, thus causing data loss, or lost or stolen, thus compromising data security. The data transmission can happen in near real-time, so that data issues could be resolved while the field team is still at the location. Data captured by this technology is usually linked to a geographical location by use of a Global Positioning System (GPS) module incorporated into modern digital devices. The GPS feature enables households to be easily identified and revisited for monitoring or validation/verification purposes. In addition, the ability of smartphones and tablets to take photographs can be used to link specific circumstances (e.g. to document an undesirable condition of the appliance) to a household.

9.1.2. Remote survey

48. For a remote survey where a surveyor will not physically visit households, data are collected through: (i) a data sensor to transmit data including figures and/or images recorded by camera; (ii) e-mail or web-based platform or SMS; (iii) telephone interview; (vi) mail (post).
49. Data sensor:
- (a) The data sensor will log and transmit data including figures and/or images recorded by camera (e.g. data logger to record the utilization hours of appliances). While the data sensor is accurate, it may be used only under certain

circumstances, and is usually expensive. For example, a pay-as-you-go system can potentially make the remote data collection more affordable as the data collected will also be the basis for monitoring lease rental payments in that case. A pay-as-you-go system is one in which one pays for a service before using it and one cannot use more than what has been paid for (e.g. pay-as-you go solar home systems).

50. Telephone interview:

- (a) Respondents require a telephone (mobile or landline) and a reliable telephone network to enable a telephone interview to take place. Additionally, there would be immediate bias problems if the statistical design of the survey relied on random sampling of telephone numbers, because households with no telephone will be excluded from the survey;
- (b) The telephone may not be suitable for long interviews, long or complex questions, and the recording of technical data that may require training.

51. E-mail or web-based platform or SMS:

- (a) The use of IT tools for data collection requires access to a computer or equivalent portal device, and a reliable Internet connection. Importantly, it may not be possible to have a statistical design where households are randomly sampled using a list of Internet users as such a list may not exist. Even if such a list exists, its storage and access may contravene local data protection laws;
- (b) If the households have access to a computer (or alternative device), collecting information via e-mail presents a number of practical problems that need careful consideration: (i) Do households regularly check their email? (ii) Different households may have different IT systems with different capabilities for displaying a questionnaire, which is particularly relevant to a questionnaire embedded in an e-mail; (iii) e-mails may be easily ignored or missed due to automatic spam filters;
- (c) Questionnaires should be relatively short with a simple structure and clear unambiguous instructions. And, importantly, the respondent must be able to understand and respond to the questionnaire;
- (d) Web-based surveys that utilize a hyperlink may be more aesthetically appealing and simpler to follow, but the practical problems above still persist. Of particular concern, whatever the communication method, is a low response rate.

52. Mailing (post) (questionnaires sent by regular mail):

- (a) Respondents answer a questionnaire received by regular mail. This requires a higher level of literacy, and the assumption that respondents are capable of understanding and completing the questionnaire. This may be facilitated by verbal explanation at a prior visit, for example;
- (b) Questionnaire fatigue⁹ needs to be addressed, as well as ensuring a simple structure for the questionnaire. The questionnaire should contain clear unambiguous instructions, have a simple format, and be short;

⁹ As more companies survey, the willingness of people to take all these surveys plummets.

- (c) This method does not allow further probing, or clarification of answers.¹⁰ Further, it may not be clear if an appropriate person has completed the questionnaire. It is well known that response rates may be very low, giving rise to bias and precision problems.

9.2. Survey and data collection methods in relation to validation/verification purposes by DOEs

9.2.1. Recommended evaluation criteria for DOE validation of survey and data collection methods used by PPs

53. PPs may choose to use different survey and data collection methods to implement a survey design, but should provide justification for the selection. This should be scrutinized as part of the validation/verification process. The following questions and evaluation criteria serve as examples and should be utilized by DOEs to validate a survey and data collection method proposed by PPs:

- (a) Why did the PPs choose the method used?
 - (i) Is the selected method compliant with the requirements of the CDM methodology?
- (b) Does the proposed data collection method match the available sampling frame?¹¹
 - (i) The appropriateness of a sampling frame will depend on the selected data collection method. Conversely, the selection of a data collection method may be influenced by which type of information is collected about the participating households. For example, a telephone survey requires a sampling frame that accurately records telephone numbers of eligible households. If such a sampling frame does not exist, is the process of its creation and maintenance included in the plan?
 - (ii) What measures are in place to ensure that non-participating households are excluded from survey and data collection methods that do not rely on physical on-site visits?
 - (iii) What mechanisms are in place to ensure that the intended recipient of the survey is the same person who completes the questionnaire? This is relevant to all survey and data collection methods;
- (c) Is the proposed survey and data collection method approach clear and suitable?
 - (i) Is there a mechanism for ensuring that the data collected are of high quality? For example during a telephone interview, the interviewer relies on the respondent giving an accurate answer to the question that is being asked. Have these mechanisms been tested in pilot telephone interviews?

¹⁰ To address this problem, it may be supported with a follow-up phone call, if available.

¹¹ A sampling frame is a complete listing of all individual units (elements, members) that can be considered as a representation of the whole population, and which can be used as a basis for selecting a sample, such as a list of all households in an area that have had solar cookers installed.

- (ii) Does the chosen data collection method suit the capability of the intended recipients? For example, a mail-based questionnaire method would be unsuitable for a target population with a low literacy rate;
- (d) Is the response rate necessary to match the required sample size reasonably achievable within the stated timelines of the data collection period? Is the stated anticipated response rate reasonable for the selected survey and data collection method?
 - (i) Is the planning information described above contained in the data collection plan? This is essential, as some methods afford weak control over the achievable response rate;
 - (ii) Is the anticipated response rate too low to match the number of required valid returns?
- (e) Is the selected survey and data collection method likely to yield results that are representative of the entire target population?
 - (i) Some survey and data collection methods (e.g. web-based surveys) are known to suffer from respondent self-selection, so yielding results that are not representative of the entire population. Is a mechanism for redressing the bias proposed? If so, is it clearly explained and supported by existing endorsed methods?
 - (ii) Does the data collection plan indicate that the existing sampling frame is fit for the intended purpose? For example, a sampling frame with telephone numbers of many digits is prone to recording errors, thus excluding eligible households whose telephone number is incorrect. What mechanisms are in place to maximize the accuracy of the sampling frame?
- (f) Is the survey and data collection method likely to provide reliable data given the nature of the parameters of interest or is it subject to measurement errors by its very nature?
 - (i) Are there questions whose answer could be subject to respondent error due to the delivery mechanism of the data collection method itself? For example, the answer “forty” units as opposed to “fourteen” units, in a telephone interview when the respondent is asked to read a meter;
 - (ii) Some data collection methods are known to suffer from under-coverage, which occurs when sections of the target population do not appear in the sampling frame. For example, do all eligible households have reliable access to the Internet? Is a mechanism for mitigating the effect of under-coverage proposed? If so, is it clearly explained and supported by existing endorsed methods?
- (g) Are the procedures for the selected survey and data collection method unambiguously defined and do they adequately provide for minimizing non-sampling errors?
 - (i) Is the quality control and assurance strategy adequate? For instance, given the common use of mobile phones, does the plan establish whether a

respondent has to be physically inside the eligible household to be able to accurately respond to a telephone interview?

- (ii) Have potential sources of bias inherent in the selected data collection method, such as self-selection and under-coverage, been anticipated? Have mechanisms for mitigating these been considered?
- (h) Does the proposed data collection plan contain the information necessary to implement the selected survey and data collection method?
 - (i) Are the proposed skill sets, qualifications and experience of the personnel engaged to conduct the data collection exercise adequate?
 - (ii) Is adequate training of personnel addressed?

9.2.2. Survey and data collection methods for use by DOEs for validation/verification of sample records of PPs

- 54. DOEs may use acceptance sampling where necessary, as described in the “Standard for sampling and surveys for CDM project activities and programme of activities”, in the validation/verification process to verify that the project proponents have implemented a sampling plan to a satisfactory standard. This involves selecting a random sample of the project proponents’ sample records and cross-checking against DOE records, i.e. data collected by the DOE. In general, different methodologies may be employed to collect survey data as described above.
- 55. The physical on-site visit, or inspection, utilizing a face-to-face interview has a clear advantage in many cases, however for simple parameters such as retention rates monitored through data sensors or pay-as-you-go systems, value addition from a site visit may be limited.
- 56. If a less preferred method such as telephone interviewing is to be conducted by a DOE this should be justified as reasonable, taking precautions to minimize bias, non-response and its effects, and implementing appropriate quality control procedures:
 - (a) Respondents require a telephone and a reliable telephone network to enable a telephone interview to take place. Bias problems would arise if not all households have access to a telephone, as the validation of corresponding records would not be possible;
 - (b) The telephone is not suitable for long interviews, long or complex questions, and the recording of technical data that may require training. Further, in telephone interviewing, it may not be clear who the respondent is, bringing into question the reliability of any answers.

9.3. Selecting a survey and data collection method and preference for use

- 57. PPs and DOEs may select one of the survey and data collection methods described above. Reasonable justification (e.g. cost-benefits) should be documented taking precautions to minimize bias, non-response and its effects, and appropriate quality control procedures. For example, if a telephone interview method is selected, it has to be justified why other preferable methods (e.g. physical visits) were not the first choice and before the telephone interview is carried out it has to be assured that all the users have

access to a telephone line. This is to guarantee that randomly selected users will be approachable through the same survey method in an unbiased, statistically correct manner.

58. If a less preferred survey and data collection method is selected then there is a requirement to justify and explain how to mitigate the potential for lower quality data, including but not limited to non-response and bias.
59. There may be strong practical reasons for using a less preferred survey and data collection method, such as security issues within a country or region.¹² Before discounting on-site visits, due consideration should be given to the use of a local third party for conducting visits. Other data collection methods may be selected, but the paragraph just above still applies.
60. If there is a possibility that planned on-site visits may not be practicable in the future due to, for example, a deteriorating security situation, PPs and DOEs should consider contingency measures at the planning stage of a survey.

9.4. Data processing and report generation

9.4.1. Data processing

61. The method used for physically collecting the data should be documented in the PDD and monitoring reports, along with a protocol for producing the final dataset.
62. If smartphone or tablet app modules are used, data can be transferred to cloud data servers (e.g. through data networks or wireless Internet on/off-line) instead of manually having the hardcopies inserted into a data sheet by other personnel. This is a cost-effective solution for data storage and data management that is being practised.
63. The protocol will list the processes employed to ensure data quality is maintained to a high standard. This should include:
 - (a) Training and experience of staff;
 - (b) Level of supervision and guidance provided to staff, for example written instructions;
 - (c) The standardized system used to initially enter data into a computer and produce a final dataset, documenting steps taken to minimize the introduction of errors. For example, the latter might be double data entry for paper-based questionnaires or application of conditional skips when using CAPI;
 - (d) Systems in place to ensure all collected data is processed;

¹² For validation purposes, paragraph 29 of the sampling standard (CDM-EB74-A06) allows flexibility in the application of the sampling standard under specific situations including security conditions. In one specific case, the Board agreed that the DOE may, for a period of one year, choose either one of the two approaches in accordance with table 1 of "Sample size and acceptance number based on AQL, UQL, and producer and consumer risk": (a) take sample size of 18, with acceptance number of 1 from the zone not affected by the conflict or (b) take sample size of 8, with acceptance number of 0 covering the entire project region. The Board also encouraged the DOE to use complementary measures (telephone calls, photographic evidence, etc.) to cross-verify the results.

- (e) The type of quality checks performed on data entered, for example range checks, inconsistency checks, checking of subsamples of data by supervisors;
- (f) How errors in (d) were recorded, reported and the remedial action taken;
- (g) Level of security and type of backup processes to guarantee data integrity, for example methods to prevent fraud and accidental deletion.

9.4.2. Report generation

64. The statistical validity of the report is judged against a number of factors. The report should contain explicit statements, explanations or content related to the following:
- (a) A suitable qualification of the data analyst;
 - (b) The statistical design of the survey, for example cluster sampling, justification of its choice and the sample size used;
 - (c) The statistical methods employed to analyse the data and reasons for choosing these methods, which should cite official or industry guidelines or other references where appropriate.

Appendix 1. Best-practice examples for sample size calculations

1. Introductory notes on sample size calculations

1. There are different equations to calculate a required sample size for different situations. Most of the examples in this document are for finite populations such as cookstoves or compact fluorescent lamps (CFLs), but there is also one example for a wastewater treatment plant where the measurements are done for the continuous flow of wastewater.
2. Which equation to use depends on the following:
 - (a) Parameter of interest, for example:
 - (i) A percentage, such as the proportion of annually operating cookstoves;
 - (ii) A numeric value, such as the mean value of operating hours of CFLs, the mean value of dry compressive strength (to check whether the manufactured bricks are of a certain quality).
3. There are other parameters, that is ratios, but this guide only covers proportions and means:
 - (a) Sampling scheme. This document contains the equations and examples using the following five sampling schemes:
 - (i) Simple random sampling;
 - (ii) Systematic sampling;
 - (iii) Stratified random sampling;
 - (iv) Cluster sampling;
 - (v) Multi-stage sampling.
4. There are a number of factors that affect the sample size required, and these are described below:
 - (a) The value that the parameter is expected to take, for example:
 - (i) Sampling to see whether 80% of installed cookstoves are still in operation will give a different sample size required than 65%. The same is true for mean values;
 - (b) The amount of variation affects the sample size required. The larger the variation associated with the parameter of interest the larger the sample size required for the same level of confidence and precision;
 - (c) The level of precision (e.g. $\pm 10\%$ of relative value of the parameter's true value) and confidence (e.g. 90% or 95%) in that precision which is desired for

determining the parameter also determines the sample size. The higher the required confidence and the narrower the precision the more samples are required.

5. Estimates of the parameter of interest (proportion, mean and standard deviation) are required for sample size calculations. There are different ways to obtain these:
 - (a) We may refer to the result of previous studies and use these results;
 - (b) In a situation where we do not have any information from previous studies, we could take a preliminary sample as a pilot and use that sample to provide our estimates;
 - (c) We could use “best guesses” based on the researcher’s own experiences.
6. Note that if the standard deviation is unknown but the range (maximum – minimum) is known then a rough “rule of thumb” is that the standard deviation can be estimated as the range divided by 4.
7. Also, for different sampling schemes additional information is required, such as strata estimates rather than just the population information.
8. There are three additional points to make in relation to these sample size calculations:
 - (a) If sample size calculations are being performed manually, it is important to retain as many decimal places as relevant, until the final calculated figure is reached. Only then should rounding be carried out. In this document, however, for clarity of presentation the detailed calculations are shown with only a small number of decimal places, although the actual calculations themselves used more than is shown;
 - (b) Researchers are encouraged to carry out more than one sample size calculation. It is highly unlikely that accurate estimates of the parameters will be available, and so the calculation should be performed for a range of possible estimates (e.g. proportion, or mean and standard deviation), and the largest sample size chosen. This should help to ensure that the sample selected will meet the required reliability criteria;
 - (c) The pilot studies that are included in the examples here are deliberately small so that calculations can be illustrated fairly easily. In real-life situations they should be larger than those used here.

2. Sample size calculations - Small-scale examples

9. For all of the small-scale examples below, we require 90% confidence that the margin of error in our estimate is not more than $\pm 10\%$ in relative terms.

2.1. Proportional parameter of interest (Cookstove project)

10. This section covers sample size calculations based on a proportion (or percentage) of interest being the objective of the project, under four different sampling schemes. Regardless of the sampling scheme used, the following have to be pre-determined in order to estimate the sample size:

- (a) The value that the proportion is expected to take;
- (b) The level of precision, and confidence in that precision (90/10 for all small-scale examples).

11. For all of the cookstove examples below, the proportion of interest is the number of project cookstoves that are still in operation at the end of the third year after the stoves were distributed; it is thought that this proportion is 0.5 (50%). The cookstoves were distributed to 640,000 households, and it has been assumed that 1 household = 1 cookstove.

2.1.1. Example 1 – Simple random sampling

12. Suppose that the population is homogeneous with respect to the continued use of the cookstoves. Then simple random sampling would be an appropriate method to estimate the proportion of cookstoves still in operation.

The equation to give us the required sample size is:

$$n \geq \frac{1.645^2 N \times p(1-p)}{(N-1) \times 0.1^2 \times p^2 + 1.645^2 p(1-p)} \quad \text{Equation (1)}$$

Where:

n	=	Sample size
N	=	Total number of households (640,000)
p	=	Our expected proportion (0.50)
1.645	=	Represents the 90% confidence required
0.1	=	Represents the 10% relative precision ($0.1 \times 0.5 = 0.05 = 5\%$ points either side of p)

13. Substituting in our values gives:

$$n \geq \frac{1.645^2 \times 640,000 \times 0.5 \times 0.5}{(640,000-1) \times 0.1^2 \times 0.5^2 + 1.645^2 \times 0.5 \times 0.5} = 270.4 \quad \text{Equation (2)}$$

14. Therefore the required sample size is at least 271 households. This assumes that 50% of the cookstoves would be operating. If we changed our prior belief of the underlying true percentage of working stoves p , this sample size would need recalculating.

- 15. Note that the figure of 271 households means 271 households with data for analysis. If we expected the response rate from the sampled households to be only 80% then we would need to scale up this number accordingly. Thus we would decide to sample $271/0.8 = 339$ households.
- 16. If we did not scale up our sample size and experienced a response rate of 80% then we would only have 216 ($271 \times 0.8 = 216$) households/cookers with data, and consequently the level of precision would be detrimentally affected. We can calculate the actual level of precision by substituting $n = 216$ into the equation:

$$\frac{1.645^2 \times 640,000 \times 0.5 \times 0.5}{(640,000 - 1) \times \textit{precision}^2 \times 0.5^2 + 1.645^2 \times 0.5 \times 0.5} = 216 \quad \text{Equation (3)}$$

- 17. This gives a relative precision of 0.1119, or 11.2% – not the 10% required. So by not adjusting our estimated sample size to take into account the expected response rate we have an increased margin of error.
- 18. One solution to this could be to take an additional sample of households. This additional sample would need to recruit 69 households which would then, again assuming a response rate of 80%, provide data on 55 households (80% of the 69). Adding these to the existing 216 gives us data for 271 households, the number required to achieve 90/10 reliability.

2.1.2. Approximate equation

- 19. The equation used above is the exact equation derived from simple random sampling theory. When population sizes are large (or infinite), then an approximate equation can be used, which ignores the actual size of the population (N). The approximate equation for the 90/10 confidence/precision guideline is:

Approximate Equation	Sample size for the above example
Proportion data $n = \frac{1.645^2(1 - p)}{0.1^2 \times p}$	$271 \left(= \frac{1.645^2 \times (1 - 0.5)}{0.1^2 \times 0.5} \right)$

2.1.2.1. Notes on approximate equations

- 20. As the sample size in this example is large, there is no difference between the sample sizes derived from the exact and approximate equations. However, for smaller populations ($N < 5000$) and small p 's (less than 0.5) there will be a difference.
- 21. Since the exact equation can be easily calculated, it is recommended that the exact equation be used in preference to the approximate one. It avoids having to decide whether the population size is large enough for it to be possible to use the approximate equation.
- 22. The scaling-up of the sample size due to non-response will also apply to the approximate equation.

2.1.3. Example 2 – Stratified random sampling

23. This time we know that stoves were distributed in four different districts and that the cookstoves are more likely to be still in operation in certain districts compared to others.¹ In this situation we want to take our knowledge about the district differences into account when we do the sampling, and sample separately from each district. Estimates of the proportion of cookstoves still in operation in each district, as well as the population size of each district are required.

District	Number of households with cookstove in district* (<i>g</i>)	Proportion of cookstoves still in operation in district (<i>p</i>)
A	76 021	0.20
B	286 541	0.46
C	103 668	0.57
D	173 770	0.33

* Note that the districts cover all of the population (sum of district populations = total population)

24. The equation for the total sample size is:

$$n \geq \frac{1.645^2 NV}{(N-1) \times 0.1^2 + 1.645^2 V} \quad \text{Equation (4)}$$

Where: $V = \frac{SD^2}{\bar{p}^2} = \frac{\text{overall variance}}{\bar{p}^2}$ and \bar{p} is the overall proportion.

25. To then decide on the number of households in the sample that come from each district we could use proportional allocation, where the proportions of units from the different districts in the sample are the same as the proportions in the population. This gives

$$n_i = \frac{g_i}{N} \times n \quad \text{where } i=1, \dots, k \quad \text{and } k \text{ is the number of districts in the area (in this case 4).}$$

Where:

g_i = Size of the i^{th} group (district) where $i=1, \dots, k$
 N = Population total

26. We use the figures from the table above to calculate the overall variance,² and proportion of cookstoves still in operation.

$$SD^2 = \frac{(g_a \times p_a(1 - p_a)) + p_b(g_b \times (1 - p_b)) + (g_c \times p_c(1 - p_c)) + \dots + (g_k \times p_k(1 - p_k))}{N} \quad \text{Equation (5)}$$

¹ If the proportions were expected to be the same in each district then simple random sampling should be used.

² The variance of a proportion is calculated as: $p(1-p)$.

$$\bar{p} = \frac{(g_a \times p_a) + (g_b \times p_b) + (g_c \times p_c) + \dots + (g_k \times p_k)}{N} \quad \text{Equation (6)}$$

Where g_i and N are as above and p_i is the proportion for the i^{th} group (district); $i=1, \dots, k$

27. Substituting the values from the table into the above equations for SD^2 and \bar{p} gives:

$$SD^2 = \frac{(76021 \times 0.20 \times 0.8) + \dots + (173770 \times 0.33 \times 0.66)}{640000} = 0.23 \quad \text{Equation (7)}$$

$$\begin{aligned} \bar{p} &= \frac{(76021 \times 0.20) + (286541 \times 0.46) + (103668 \times 0.57) + (173770 \times 0.33)}{640000} \quad \text{Equation (8)} \\ &= 0.41 \end{aligned}$$

Therefore:

$$V = \frac{SD^2}{\bar{p}^2} = \frac{0.23}{0.41^2} = 1.37 \quad \text{Equation (9)}$$

28. Substituting in V into our sample size equation gives:

$$n \geq \frac{1.645^2 \times 640000 \times 1.37}{(640000 - 1) \times 0.1^2 + 1.645^2 \times 1.37} = 367.0 \quad \text{Equation (10)}$$

29. The total sample size required is 367 households. This then needs to be divided up according to the size of each district to get the number of households that should be sampled in each district.

$$\text{General equation } n_i = \frac{g_i}{N} \times n \quad \text{Equation (11)}$$

$$\text{District A: } n_a = \frac{76021}{640000} \times 367 = 43.7$$

$$\text{District B: } n_b = \frac{286541}{640000} \times 367 = 164.8$$

$$\text{District C: } n_c = \frac{103668}{640000} \times 367 = 59.6$$

$$\text{District D: } n_d = \frac{173770}{640000} \times 367 = 99.9$$

30. Rounding up the district samples sizes gives the number of households to be sampled in each district, 44 in A, 165 in B, 60 in C, and 100 in D (the sum of these is slightly greater than the total required sample size due to the rounding up of households within each district).

31. Note that these sample sizes do not take into account non-response. If the expected level of response is 75% across all districts then divide each district sample size by 0.75; this will result in larger sample sizes allowing for the non-responders.

2.1.4. Example 3 – Cluster sampling

32. Now consider a different scenario. The households are not located in different districts. Instead they are ‘clustered’ or grouped into lots of villages. Instead of going to numerous individual households, we want to go to a number of villages and sample every household within each village.
33. For this example the population comprises 120 villages, all of approximately similar size. In order to have some understanding of the proportion of cookstoves still operating and the variation in this proportion between villages, a small preliminary sample has been taken:

Village	Estimated proportion of cookstoves operating in each village
1	0.37
2	0.48
3	0.50
4	0.27
5	0.68
Average (\bar{p})	0.46
Variance SD_B^2	0.024

34. The average (\bar{p}) is just $\frac{0.37 + 0.48 + 0.50 + 0.27 + 0.68}{5} = \frac{2.3}{5} = 0.46$ and the variance between the clusters is:

$$SD_B^2 = \sum_{i=1}^{n=5} ((p_i - \bar{p}))^2 \quad \text{Equation (12)}$$

$$= \frac{(0.37 - 0.46)^2 + (0.48 - 0.46)^2 + \dots + (0.68 - 0.46)^2}{4}$$

$$= \frac{0.0946}{4} = 0.0237$$

35. The equation for the number of villages that need to be sampled is:

$$c \geq \frac{1.645^2 MV}{(M-1) \times 0.1^2 + 1.645^2 V} \quad \text{Equation (13)}$$

Where:

$$V = \frac{SD_B^2}{\bar{p}^2} = \frac{\text{variance between clusters (villages)}}{\text{average proportion}}$$

C = Number of clusters to be sampled (villages)

M = Total number of clusters (villages) – this must encompass the entire population

1.645 = Represents the 90% confidence required

0.1 = Represents the 10% relative precision required

36. Substituting our values into the above equation gives the number of villages that are required to be sampled as

$$V = \frac{SD_B^2}{\bar{p}^2} = \frac{0.0237}{0.46^2} = 0.11 \quad \text{Equation (14)}$$

$$c \geq \frac{1.645^2 \times 120 \times 0.11}{(120-1) \times 0.1^2 + 1.645^2 \times 0.11} = 24.3 \quad \text{Equation (15)}$$

37. Therefore we would have to sample every household within 25 randomly selected villages. This approach to sampling assumes that the villages are homogeneous. In this example this means that the proportion of cookstoves still operating in a village is independent of any other factors such as district (see example 2 – stratified sampling), economic status, etc. If the proportions are not independent of another factor then cluster sampling within each strata of the factor can be used.

38. Since cluster sampling is dealing with data from whole clusters (villages in this example), non-response at the within-village level (household in this case) is less likely to be an issue, unless there is a high percentage of non-responses within a village. If there are only one or two missing values in a village it is still possible to obtain a usable proportion for that village based on all the other households that did provide data.

2.1.5. Example 4 – Multi-stage sampling

39. Multi-stage sampling can be thought of as sampling from a number of groups, and then going on to sample units within each group. Continuing with the cookstove example, we want to sample a number of villages and then a number of households within each sampled village.

40. We know that there are 120 villages and there are on average 50 households within each village, of which we plan to sample 10. From a small pilot study we already know the following:

Village	Proportion of cookstoves in operation
A	0.37
B	0.48
C	0.50
D	0.27
E	0.68

41. The equation for the number of villages to be sampled is:

$$c \geq \frac{\frac{SD_B^2}{p^2} \times \frac{M}{M-1} + \frac{1}{u} \times \frac{SD_w^2}{p^2} \times \frac{(\bar{N} - \bar{u})}{(\bar{N} - 1)}}{\frac{0.1^2}{1.645^2} + \frac{1}{M-1} \frac{SD_B^2}{p^2}} \tag{Equation (16)}$$

Where:

- C = Number of groups that should be sampled
- M = Total number of groups in the population (120 villages)
- \bar{u} = Number of units to be sampled within each group (pre-specified as 10 households)
- \bar{N} = Average units per group (50 households per village)
- SD_B^2 = Unit variance (variance between villages)
- SD_w^2 = Average of the group variances (average within village variation)
- p = Overall proportion
- 1.645 = Represents the 90% confidence required
- 0.1 = Represents the 10% relative precision

42. Using our table of pilot information we can calculate the unknown quantities for the equation above.

Village	Proportion of cookstoves in operation (p_i)	Variance within village ($p_i(1-p_i)$)
A	0.37	0.2331
B	0.48	0.2496
C	0.50	0.2500
D	0.27	0.1971

Village	Proportion of cookstoves in operation (p_i)	Variance within village ($p_i(1-p_i)$)
E	0.68	0.2176
Average	$\bar{p} = 0.46$	$SD_W^2 = 0.2295$
Variance	$SD_B^2 = 0.0237$	

Where:

\bar{p} = The average proportion of cookstoves, i.e.

$$\frac{0.37 + \dots + 0.68}{5} = 0.46$$

SD_W^2 = The average variance within the villages, i.e.

$$SD_W^2 = \frac{0.2331 + \dots + 0.2176}{5} = 0.2295$$

SD_B^2 = The variance between the village proportions, i.e. the variance between 0.37, 0.48 etc. This can be calculated in the usual way for calculating a variance i.e. using the equation

$$SD_B^2 = \frac{\sum_{i=1}^n (p_i - \bar{p})^2}{n-1}$$
 which gives $SD_B^2 = 0.0237$

43. Substituting our values into the group sample size equation gives:

$$c \geq \frac{\frac{0.0237}{0.46^2} \times \frac{120}{(120-1)} + \frac{1}{10} \times \frac{0.2295}{0.46^2} \times \frac{(50-10)}{(50-1)}}{\frac{0.1^2}{1.645^2} + \left(\frac{1}{(120-1)} \times \frac{0.0237}{0.46^2} \right)} = 43.4$$

Equation (17)

44. Therefore if we were to sample 10 households from each village we should sample 44 villages for the required confidence/precision.

45. It is usually useful to have this calculation automated so that a series of different u values (the number of units to be sampled in each group) can be used and the effect that this has on the number of groups to be sampled can be observed.

Number of households sampled in each village u	Required number of villages c
5	68
10	44
15	36
20	32
30	28
50	25

46. In this example, by doubling the number of households within each village to be sampled from 10 to 20, we reduce the number of villages that need to be visited from 44 to 32.
47. Note that when u = the average number of households in a village (50), the required sample size is the same as that from cluster sampling as everyone within each village would be sampled. When u is smaller than the average number of households, the number of villages that need to be sampled under multi-stage sampling is greater than that from cluster sampling as not everyone within each village is being sampled.

2.1.6. Mean value parameter of interest (CFL project)

48. This section covers sample size calculations where the objective of the project relates to a mean value of interest, under four different sampling schemes. For the sample size calculations, regardless of the sampling scheme, we need to know:
- (a) The expected mean (the desired reliability is expressed in relative terms to the mean);
 - (b) The standard deviation;
 - (c) The level of precision, and confidence in that precision (90/10 for all small-scale examples).
49. The examples below are based on the parameter of interest being average daily CFL usage, which is thought to be 3.5 hours, with a standard deviation of 2.5 hours. The population consists of 420,000 households to which CFLs were distributed; we are assuming that 1 household = 1 CFL.

2.1.7. Example 5 – Simple random sampling

50. For simple random sampling to be appropriate we are assuming that CFL usage is homogeneous amongst the households.
51. The following equation can be used to calculate the sample size:

$$n \geq \frac{1.645^2 NV}{(N-1) \times 0.1^2 + 1.645^2 V} \quad \text{Equation (18)}$$

Where:

$$V = \left(\frac{SD}{mean} \right)^2$$

n = Sample size

N = Total number of households

$Mean$ = Our expected mean (3.5 hours)

SD = Our expected standard deviation (2.5 hours)

1.645 = Represents the 90% confidence required

0.1 = Represents the 10% relative precision

$$V = \left(\frac{2.5}{3.5}\right)^2 = 0.51 \tag{Equation (19)}$$

$$n = \frac{1.645^2 \times 420,000 \times 0.51}{(420,000 - 1) \times 0.1^2 + 1.645^2 \times 0.51} = 138.0 \tag{Equation (20)}$$

- 52. Therefore the required sample size is at least 138 households.
- 53. Note that if we expected the response rate from the sampled households to be only 70% then we would need to scale up the number obtained above accordingly. Thus we would decide to sample $138/0.7 = 198$ households.

2.1.8. Approximate equation

- 54. The equation used above is the exact equation. When population sizes are large (or infinite), then approximate equations can be used, which ignore the actual size of the population (N).
- 55. The approximate equation follows the 90/10 confidence/precision guideline:

Approximate Equation	Sample size for the above example
Mean value data $n = \frac{1.645^2 V}{0.1^2}$ Where: $V = \left(\frac{SD}{mean}\right)^2$	$138 \left(= \frac{1.645^2 \times 0.51}{0.1^2} \right)$

- 56. Please see the “Approximate equation” section under **i) Cookstove Project - Proportional parameter of interest, Example 1: Simple random sampling** for notes relating to approximate equations.

2.1.9. Example 6 – Stratified random sampling

- 57. The key to this example is that, unlike under simple random sampling, it is not assumed that the population is homogeneous – different parts of the population are expected to have different CFL usage averages.
- 58. Suppose that the CFLs were distributed in different districts in which each has a different CFL usage pattern (due to district economic backgrounds). We are now interested in sampling users of CFLs from all the districts to ensure all areas are well represented.

59. Each district has the following number of households, and mean and standard deviation CFL usage:

District	Number of households in district given a CFL	Mean (hours)	Standard deviation (hours)
A	146 050	3.2	1.9
B	104 474	2.4	0.8
C	38 239	4.5	1.6
D	74 248	1.6	1.7
E	56 989	2.3	0.7

60. The total sample size of households across all five districts is:

$$n \geq \frac{1.645^2 \times NV}{(N-1) \times 0.1^2 + 1.645^2 V} \tag{Equation (21)}$$

Where:

$$V = \left(\frac{SD}{mean} \right)^2$$

SD = Is the overall standard deviation, and

Mean = Is the overall mean

61. Using the data in the table above we can estimate the overall mean and standard deviation. Both equations are weighted according to the total number of households in each district.
62. Overall Standard Deviation:

$$SD = \sqrt{\frac{(g_a \times SD_a^2) + (g_b \times SD_b^2) + (g_c \times SD_c^2) + \dots + (g_k \times SD_k^2)}{N}} \tag{Equation (22)}$$

Where:

SD = Weighted overall standard deviation
SD_i Standard deviation of the *i*th group where *i*=1,...,k, (note that these are all squared – so the group size is actually being multiplied by the group variance)

g_a = Size of the *i*th group where *i*=1,...,k

N = Population total

$$\text{mean} = \frac{(g_a \times m_a) + (g_b \times m_b) + (g_c \times m_c) + \dots + (g_k \times m_k)}{N} \quad \text{Equation (23)}$$

Where:

mean = Weighted overall mean

m_i = Mean of the i^{th} group where $i=1, \dots, k$

63. Substituting the values from our example into the above expressions gives:

$$SD = \sqrt{\frac{(146050 \times 1.9^2) + (104474 \times 0.8^2) + \dots + (56989 \times 0.7^2)}{420000}} = 1.49 \quad \text{Equation (24)}$$

$$\text{mean} = \frac{(146050 \times 3.2) + \dots + (56989 \times 2.3)}{420000} = 2.71 \quad \text{Equation (25)}$$

64. Substituting these values into the equation for V gives:

$$V = \left(\frac{SD}{\text{mean}} \right)^2 = \left(\frac{1.49}{2.71} \right)^2 = 0.3 \quad \text{Equation (26)}$$

65. And hence, for the sample size:

$$n = \frac{1.645^2 \times 420,000 \times 0.3}{(420,000 - 1) \times 0.1^2 + 1.645^2 \times 0.3} = 81.7 \quad \text{Equation (27)}$$

66. This example assumes proportional allocation, which means that the number of households we want to sample from each district is proportional to the size of the district within the population.

The equation for each district sample size is: $n_i = \frac{g_i}{N} \times n$ Equation (28)

$$\text{District A: } n_a = \frac{146050}{420000} \times 82 = 29$$

$$\text{District B: } n_b = \frac{104474}{420000} \times 82 = 21$$

$$\text{District C: } n_c = \frac{38239}{420000} \times 82 = 8$$

$$\text{District D: } n_d = \frac{74248}{420000} \times 82 = 15$$

$$\text{District E: } n_e = \frac{56989}{420000} \times 82 = 12$$

- 67. The summation of these district sample sizes (29+21+8+15+12=85) is slightly greater than that calculated from the total sample size equation (82) above due to rounding.
- 68. As with previous examples, the sample sizes above need to be scaled up to take into account any non-response expected.

2.1.10. Example 7 – Cluster sampling

- 69. Suppose CFLs were distributed to households in 50 villages. Instead of sampling from the whole population of households with CFLs, we sample a number of villages (villages=clusters), and then collect data from all households within the villages.
- 70. The equation used to give us the required number of clusters, *c*, to sample is:

$$c \geq \frac{1.645^2 MV}{(M-1) \times 0.1^2 + 1.645^2 V} \tag{Equation (29)}$$

Where:

- V* = $\left(\frac{SD}{Cluster\ mean}\right)^2$
- M* = Total number of clusters (50 villages)
- 1.645 = Represents the 90% confidence required
- 0.1 = Required precision

- 71. To perform the calculations we need information about CFL usage at the village level, rather than at the household level. If such information does not already exist, we could possibly collect it in a pilot study. The example here assumes that data are available from a pilot study on five villages.

Village	Total usage across all households in the village ³
A	30 458
B	27 667
C	31 500
D	28 350
E	19 125

³ In the pilot study these totals may be derived from collecting data on all households in the village, or else by taking a sample of households in the village and scaling up from the sample to all households.

72. Calculating the mean and standard deviation of these figures gives:

$$\text{Cluster mean } (\bar{y}) = \frac{1}{n} \sum_{i=1}^n y_i = \frac{30458 + 27667 + \dots + 19125}{5} = 27420 \quad \text{Equation (30)}$$

$$SD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{where the } y_i \text{ are the total usages for the villages.}$$

$$SD_B^2 = 23902660 \quad \text{and so } SD_B = 4889$$

73. These statistics (i.e. mean and SD of data) are easily produced using statistical software.
74. Substituting these values into the equation gives the required number of clusters, i.e. villages as:

$$V = \left(\frac{4889}{27420} \right)^2 = 0.03 \quad \text{Equation (31)}$$

$$c \geq \frac{1.645^2 \times 50 \times 0.03}{(50-1) \times 0.1^2 + 1.645^2 \times 0.03} = 7.5 \quad \text{Equation (32)}$$

75. So we need to sample eight villages to satisfy the 90/10 confidence/precision criterion. Once a village is selected, all households in the selected village should be sampled.
76. The above equation assumes that CFL usage in a village is independent of any other factors, such as economic status. If CFL usage was expected to vary according to another factor then cluster sampling can be used within each level of the factor.

2.1.11. Example 8 – Multi-stage sampling

77. Multi-stage sampling combines the cluster and simple random sampling approaches in a two-stage sampling scheme which enables us to randomly select some groups (villages) and then randomly sample some units (households) within those groups (villages). As with simple random sampling and cluster sampling, we are assuming homogeneity across villages in the usage of CFLs. We know that the 420,000 households are in 50 villages.
78. Let us start by assuming that we want to sample 10 households in each village. In general terms we will call this number u (for units).
79. In order to perform a sample size calculation we need information on:
- The variation between households within the villages;
 - The variation between villages;
 - The average household usage;

(d) The average usage at the village level.

80. A previous study had provided data for households in five villages, and the results are summarized below. Note that not all villages in this example are exactly the same size.

CFL average daily usage (hours)				
Village	Number of households	Mean usage ⁴ per household in a village	Total usage across all households	Standard deviation ⁵ (between households within villages)
A	8 500	3.58	30 458	2.60
B	8 300	3.33	27 667	2.70
C	8 400	3.75	31 500	0.66
D	8 100	3.50	28 350	0.75
E	8 500	2.25	19 125	1.50
Total number of households	41 800			
Overall mean usage per household		3.28		
Mean usage per village			27 420	
SD _B = Standard deviation between villages (SD of the total usage column)			4 889	
SD _W = Average within village standard deviation				1.86

81. In the table above, the overall mean CFL usage is the average usage for a household, i.e. $Overall\ mean = \frac{30458 + 27667 + \dots + 19125}{41800} = 3.28$

82. The cluster or village mean CFL usage is the average usage for village, i.e. $Cluster\ mean = \frac{30458 + 27667 + \dots + 19125}{5} = 27420$

83. SD_W^2 is the average of the variances between households within the villages. Its square root (SD_W) is the average within village standard deviation. The equation for SD_W^2 is:

$$SD_W^2 = \frac{8500 \times 2.60^2 + \dots + 8500 \times 1.50^2}{41800} = 3.48 \text{ and so } SD_W = 1.86$$

SD_B^2 is the variance between the village total usages and its square root is the standard deviation between villages. It can be calculated using the usual equation for a variance, i.e.

$$SD_B^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} \text{ where the } y_i \text{ are the total usages for the villages.}$$

⁴ This can be a mean from all households or a mean from a sample of households.

⁵ And this can be a standard deviation based on all households or a sample of households.

$$SD_B^2 = 23902660 \text{ and so } SD_B = 4889$$

84. We have pre-specified that we want to sample 10 households within each village, so we need to calculate how many villages need to be sampled given a 90/10 confidence/precision criterion is required:

$$c \geq \frac{\left(\frac{SD_B}{Clustermean}\right)^2 \times \left(\frac{M}{M-1}\right) + \left(\frac{1}{u}\right) \times \left(\frac{SD_W}{Overallmean}\right)^2 \left(\frac{\bar{N}-u}{\bar{N}-1}\right)}{\left(\frac{0.1}{1.645}\right)^2 + \frac{1}{M-1} \left(\frac{SD_B}{Clustermean}\right)^2} \tag{Equation (33)}$$

Where:

- M = Total number of groups (50 villages)
- \bar{N} = Average number of units per group (approximately 8,400 households per village)
- U = Number of units that have been pre-specified to be sampled per group (pre-specified number of households to be sampled in each village = 10)
- 1.645 = Represents the 90% confidence required
- 0.1 = Required precision

$$c \geq \frac{\left(\frac{4889}{27420}\right)^2 \times \left(\frac{50}{50-1}\right) + \left(\frac{1}{10}\right) \times \left(\frac{1.86}{3.28}\right)^2 \left(\frac{8400-10}{8400-1}\right)}{\left(\frac{0.1}{1.645}\right)^2 + \frac{1}{50-1} \left(\frac{4889}{27420}\right)^2} = 14.9 \tag{Equation (34)}$$

85. Therefore if we were to sample 10 households from each village we should sample 15 villages for the required confidence/precision.
86. It is usually useful to have this calculation automated so that a series of different u values (the number of units to be sampled in each group) can be used and the effect that this has on the number of groups to be sampled can be observed.

Number of households sampled in each village u	Required number of villages c
5	23
10	15
15	13
20	12
25	11

87. Compared to the cluster sampling example, more villages are required for the multi-stage sampling scheme because fewer households are being sampled within each village.

88. Note that in the above example the villages are of slightly different sizes. In practice this is likely to be the case, although the actual sizes may not always be known. This is not critical to the sample size calculation. What is important is that sensible estimates of the mean and standard deviation at both the cluster level (village level) and unit level (household level) are used in the calculation.

2.2. Mean value parameter of interest (Brick project)

89. This section covers an example sample size calculation based on systematic sampling where the objective of the project relates to a mean value of interest.

90. As for all mean value parameters of interest examples we need to know:

- (a) The expected mean (the desired reliability is expressed in relative terms to the mean);
- (b) The standard deviation;
- (c) The level of precision, and confidence in that precision (90/10 for all small-scale examples).

91. The following example is based on assessing whether bricks are of a minimum quality after manufacture; dry compressive strength has been identified as a suitable measurement of quality. Prior information gives us a mean dry compressive strength of 158kg/cm² with a standard deviation of 65kg/cm².

2.2.1. Example 9 – Systematic sampling

92. This example is based on a manufacturing process; we want to systematically sample every nth brick from the production line of 500,000 bricks per year. We wish to know how many bricks should be sampled to ensure an average dry compressive strength of 158kg/cm², with 90/10 confidence/precision.

93. The sample size equation for a required 90/10 confidence/precision is:

$$n \geq \frac{1.645^2 V}{0.1^2} \quad \text{Equation (35)}$$

$$\text{Where: } V = \left(\frac{SD}{mean} \right)^2$$

94. Substituting in mean and standard deviation from above gives:

$$V = \left(\frac{65}{158} \right)^2 = 0.17 \quad \text{Equation (36)}$$

$$n \geq \frac{1.645^2 \times 0.17}{0.1^2} = 45.8 \quad \text{Equation (37)}$$

95. Therefore, we should take 46 samples to gain the required levels of confidence and precision. Given that 500,000 bricks are manufactured each year and we want to take 46 samples, we should sample 1 brick for every N/n bricks produced – that is 1 brick for every 10,917 ($= 500,000 / 46$) bricks.
96. To make sure our sample is random, we randomly choose a starting point (starting brick) between 1 and 10,917 and use this brick as our first sample – for example brick 6505. We continue sampling by taking every 10,917th brick, so our second sample would be brick $6505 + 10,917 = 17,422$, the third brick sampled would be $17,422 + 10,917 = 28,339$, etc. For the sake of practicality it might be easier to sample every 10,000th brick; instead of every 10,917th, this would give a slightly larger sample size.

2.3. Measurements in biogas projects

2.3.1. Example 10

97. A survey will be carried out to estimate the mean chemical oxygen demand (COD) at a wastewater plant. The wastewater is a continuous flow of water that leaves the plant. A 500 ml sample of water will be extracted (from plant inlet) from the continuous flow of wastewater on a regular basis throughout the year and a single measurement of COD (mg/L) made on each sample.
98. This form of sampling, i.e. on a regular basis, possibly with a random start date, is systematic sampling.
99. The wastewater system has been in place for some time, and is considered to be stable in terms of the way it is functioning. The COD for the inlet is thought to be at a constant level throughout the year (apart from random variation)⁶.
100. Previous work where measurements were taken on a regular basis suggested that the mean COD for untreated water is likely to be about 31,750 mg/L and the standard deviation (*SD*) in the order of 6,200 mg/L.
101. Since the wastewater is flowing continuously, the study population can be thought of as all possible 500 ml water samples in a whole year – so large as to be almost infinite. The sample size calculation no longer needs inclusion of the finite population size (i.e. *N*).
102. If the sampling times are sufficiently far apart the data can be regarded as a set of independent observations and treated as a simple random sample. The number of COD measurements that are required to meet the 90/10 reliability is:

$$n = \left(\frac{t_{n-1} \times SD}{0.1 \times \text{mean}} \right)^2 \quad \text{Equation (38)}$$

103. Where t_{n-1} is the value of the t-distribution for 90% confidence when the sample size is n .⁷ However, the sample size is not yet known, and so a first step is to use the value for 90% confidence when the sample is large, i.e. 1.645, and then refine the calculation.

⁶ In reality, temporal fluctuations (daily, weekly, seasonally, etc.) both in the wastewater flow and COD concentration should be taken into account when taking samples.

$$n = \left(\frac{1.645 \times SD}{0.1 \times mean} \right)^2 \quad \text{Equation (39)}$$

104. This gives $n = \left(\frac{1.645 \times 6200}{0.1 \times 31750} \right)^2 = 10.3$ which rounds up to 11.
105. The calculation now needs to be repeated using the t-value for 90% confidence and $n=11$.
106. The exact figure for this t-value can be acquired from any set of general statistical tables or using standard statistical software. For a sample size of 11 the value is 1.812.
107. The calculation now gives $n = \left(\frac{1.812 \times 6200}{0.1 \times 31750} \right)^2 = 12.5$ which rounds up to 13.
108. The process should be iterated until there is no change to the value of n . Here the repeat calculation would have a t-value of 1.782 and the calculation would yield $n = 12.11$, which would be rounded up to 13. The sample size calculation suggests that sampling every four weeks should be sufficient for 90/10 reliability.

2.4. Other calculated sample sizes

109. The above is a relatively simple example, and not all situations will yield values as neat as “once a month” or “once every four weeks”. For instance:
- Had the calculation indicated that 48 measurements should be taken, one would most likely decide to sample weekly for the whole year;
 - If the calculation had indicated 16 samples were required, then one might decide to sample every three weeks. Alternatively, since this may not be an easy schedule to comply with, one might choose to sample every two weeks. This now gives us a total of 26 samples which should ensure that the data, when collected and analysed, have more than adequate precision (assuming, of course, that the figures for the mean and the standard deviation that were used in the sample size calculation were good reflections of the true situation).
110. Instead of trying to follow “unworkable” schedules, it may be more sensible to use the following simplifications:

No. of measurements determined from sample size calculation	Proposed schedule
Less than or equal to 12	Monthly
13–17	Every three weeks
18–26	Every two weeks
26–51	Every week

⁷ This is indicated by the subscript ($n-1$) which is called the degrees of freedom for the t-value.

No. of measurements determined from sample size calculation	Proposed schedule
More than 52	Twice a week

2.5. Understanding variation

111. The above example illustrates the sample size calculation using an absolute figure for the standard deviation. However, sometimes researchers have difficulty providing a figure for the standard deviation, but they can express it in relative terms. For instance, when asked about the variation in COD in this wastewater example, the researcher may describe it as 20%.

112. The coefficient of variation (CV) is a summary measure which describes variability in terms of the mean. The actual equation is $CV = \frac{SD}{mean}$. It is sometimes multiplied by 100, in which case it is describing the standard deviation as a percentage of the mean. The sample size equation on the previous page can now be written as $n = \left(\frac{t_{n-1} \times CV}{0.1} \right)^2$

where t_{n-1} is the value of the t-distribution for 90% confidence for a sample of n measurements. Again the value of 1.645 would be used instead of a t-value for the first step in the calculation; and so in this example the first step would be

$$n = \left(\frac{1.645 \times 0.2}{0.1} \right)^2 = 10.8.$$

3. Sample size calculations - Large-scale examples

113. For the large-scale examples we require 95% confidence that the margin of error in our estimate is not more than $\pm 10\%$ in relative terms.

3.1. Proportional parameter of interest (Transport project)

114. This section covers sample size calculations based on a proportion (or percentage) of interest being the objective of the project, under four different sampling schemes. Regardless of the sampling scheme used, the following have to be pre-determined in order to estimate the sample size based on a proportion:

- (a) The value that the proportion is expected to take;
- (b) The level of precision, and confidence in that precision (95/10 for all large-scale examples).

115. The examples relate to passengers travelling on the transport project in Bogota. It is known that 1,498,630 passengers use the project for transportation every day; the parameter of interest is the proportion of these passengers that would previously have travelled by bus, thought to be 45%.

3.1.1. Example 11 – Simple random sampling

116. Assuming that the proportion of interest is homogeneous. The equation for the sample size required under simple random sampling is:

$$n \geq \frac{1.96^2 NV}{(N-1) \times 0.1^2 + 1.96^2 V} \quad \text{Equation (40)}$$

Where:

$$V = \frac{p(1-p)}{p^2}$$

N = Sample size with finite population correction

N = Total number of passengers per day

P = Our estimated proportion (45%)

1.96 = Represents the 95% confidence required

0.1 = Required precision

117. Substituting our values into the above equation we get:

$$V = \frac{0.45 \times (1 - 0.45)}{0.45^2} = 1.22 \quad \text{Equation (41)}$$

$$n \geq \frac{1.96^2 \times 1,498,630 \times 1.22}{(1,498,630 - 1) \times 0.1^2 + 1.96^2 \times 1.22} = 469.4 \quad \text{Equation (42)}$$

118. The required sample size is at least 470 passengers to get an estimated proportion of passengers that would have previously travelled by bus with 95/10 confidence/precision. Note that the sample size will change depending on the estimated proportion value.
119. The above sample size does not take into account any non-responders, that is passengers who do not respond to the question. If we expect that 90% of passengers will respond (and 10% will not) then we should increase the sample size by dividing the sample size calculated above by the expected level of response: $470/0.9 = 523$. To account for a non-response of 10% we should sample 523 passengers.
120. Please see the “Approximate equation” section under i) **Cookstove Project - Proportional parameter of interest, Example 1: Simple random sampling** for notes relating to approximate equations. Note that 1.96 should be used in place of 1.645 to account for the increased confidence required for the large-scale projects.

3.1.2. Example 12 – Stratified random sampling

121. Suppose that we believe the proportion of transport project passengers that would have previously used the bus would vary between the eight different zones in which the

project operates. We would like to make sure that when we do our sampling, our sample includes a representative proportion of passengers from each zone. To calculate the sample size, estimates of the number of passengers and proportion that would have previously travelled by bus within each zone are required.

Zone	Number of passengers within each zone (g)	Estimated proportion of passengers that would have used a bus (p)
A	19 865	0.43
B	21 358	0.57
C	30 1245	0.4
D	65 324	0.71
E	654 832	0.32
F	50 213	0.46
G	12 489	0.26
H	373 304	0.68

122. The equation for the total sample size is:

$$n \geq \frac{1.96^2 NV}{(N-1) \times 0.1^2 + 1.96^2 V} \tag{Equation (43)}$$

Where: $V = \frac{SD^2}{\bar{p}^2} = \frac{\text{overall variance}}{(\text{overall proportion})^2}$

123. To then decide on the number of passengers in the sample that come from each zone we could use proportional allocation, where the proportions of units from the different zones in the sample is the same as the proportions in the population. This gives:

$$n_i = \frac{g_i}{N} \times n \text{ where } i=1, \dots, k \text{ where } k \text{ is the number of zones (in this case 8).}$$

Where:

g_i = Size of the i^{th} group (district) where $i=1, \dots, k$

N = Population total

124. Using the figures from the table we can calculate the overall variance,⁸ and overall proportion:

$$SD^2 = \frac{(g_a \times p_a(1 - p_a)) + (g_b \times p_b(1 - p_b)) + (g_c \times p_c(1 - p_c)) + \dots + (g_k \times p_k(1 - p_k))}{N} \tag{Equation (44)}$$

⁸ The variance of a proportion is calculated as: $p(1-p)$.

$$\bar{p} = \frac{(g_a \times p_a) + (g_b \times p_b) + (g_c \times p_c) + \dots + (g_k \times p_k)}{N} \quad \text{Equation (45)}$$

Where: g_i and N are as above and p_i is the proportion of the i^{th} group (district) where $i=1, \dots, k$.

$$SD^2 = \frac{(19865 \times 0.43 \times 0.57) + (21358 \times 0.57 \times 0.43) + \dots + (373304 \times 0.28 \times 0.72)}{1,498,630} = 0.22 \quad \text{Equation (46)}$$

$$\bar{p} = \frac{(19865 \times 0.43) + (21358 \times 0.57) + \dots + (373304 \times 0.28)}{1,498,630} = 0.45 \quad \text{Equation (47)}$$

125. Therefore:

$$V = \frac{SD^2}{\bar{p}^2} = \frac{0.22}{0.45^2} = 1.09 \quad \text{Equation (48)}$$

126. Substituting in our V gives:

$$n \geq \frac{1.96^2 \times 1,498,630 \times 1.09}{(1,498,630 - 1) \times 0.1^2 + 1.96^2 \times 1.09} = 419.6 \quad \text{Equation (49)}$$

127. The total sample size required is 420 passengers. The next step is to divide this total sample size up according to the size of each zone to get the number of passengers to be sampled within each zone.

$$\text{General equation: } n_i = \frac{g_i}{N} \times n \quad \text{Equation (50)}$$

$$\text{Zone A: } n_a = \frac{19865}{1,498,630} \times 420 = 5.6$$

$$\text{Zone B: } n_b = \frac{21358}{1,498,630} \times 420 = 6.0$$

$$\text{Zone C: } n_c = \frac{301245}{1,498,630} \times 420 = 84.4$$

$$\text{Zone D: } n_d = \frac{65324}{1,498,630} \times 420 = 18.3$$

$$\text{Zone E: } n_e = \frac{654832}{1,498,630} \times 420 = 183.5$$

$$\text{Zone F: } n_f = \frac{50213}{1,498,630} \times 420 = 14.1$$

$$\text{Zone G: } n_g = \frac{12489}{1,498,630} \times 420 = 3.5 \qquad \text{Zone H: } n_h = \frac{373304}{1,498,630} \times 420 = 104.6$$

- 128. Rounding up the zone sample sizes gives the number of passengers to be sampled in each zone (the sum of these is slightly greater than the required sample size due to the rounding up of passengers within each zone). The sample sizes required vary so much between the zones because the number of passengers in each zone is so different.
- 129. Note that these sample sizes do not take into account non-response. If the expected level of response is 85% across all zones then divide each zone sample size by 0.85. This will result in larger sample sizes allowing for the non-responders.

3.1.3. Example 13 – Cluster sampling

- 130. Instead of sampling individual passengers, it has been decided that buses (clusters) are going to be sampled and then every passenger on each of the selected buses will be asked if they travelled by bus prior to the project. To calculate the sample size we require the number of clusters that make up the population, that is the number of buses that carry the transport project passengers; for this example we will assume 12,000 buses. We also need estimated proportions of passengers that would have travelled by bus prior to the project from a number of buses; for this example we have previously sampled four buses and the proportions were:

Bus	Estimated Proportion
1	0.37
2	0.46
3	0.28
4	0.52
Average (\bar{p})	0.4075
Variance (SD_B^2)	0.011

- 131. The equation for the number of buses that need to be sampled is:

$$c \geq \frac{1.96^2 MV}{(M-1) \times 0.1^2 + 1.96^2 V} \qquad \text{Equation (51)}$$

Where:
$$V = \frac{SD_B^2}{\bar{p}^2} = \frac{\text{Variance between clusters(buses)}}{\text{Average proportion over clusters}}$$

132. The average proportion is just $\frac{0.37 + 0.46 + 0.28 + 0.52}{4} = \frac{1.63}{4} = 0.41$ and the variance between the clusters is:

$$SD_B^2 = \frac{1}{n-1} \sum_{i=1}^{n=5} (p_i - \bar{p})^2$$

$$= \frac{(0.37 - 0.4065)^2 + (0.46 - 0.4065)^2 + \dots + (0.52 - 0.4065)^2}{3}$$

$$= 0.0110$$

Equation (52)

Where:

- c = Number of clusters to be sampled (buses)
- M = Total number of clusters (buses) - this must encompass the entire population
- 1.96 = Represents the 95% confidence required
- 0.1 = Represents the 10% relative precision

133. Substituting our values into the above equation gives:

$$V = \frac{SD_B^2}{\bar{p}^2} = \frac{0.0110}{0.41^2} = 0.07$$

Equation (53)

$$c \geq \frac{1.96^2 \times 12000 \times 0.0664}{(12000 - 1) \times 0.1^2 + 1.96^2 \times 0.0664} = 25.5$$

Equation (54)

134. Therefore we would have to sample every passenger on 26 randomly selected buses.
135. This approach to sampling assumes that the population is homogeneous. In this example this means that the proportion of passengers that would have previously travelled by bus is independent of any other factors such as zones (see example 12 – stratified sampling), economic status, etc. If the proportion of passengers that would have previously travelled by bus is expected to be different for different zones, then cluster sampling should be used within each zone.
136. Non-response is unlikely to be a problem when using cluster sampling, unless the number of individuals within a cluster could be 0 (a bus with no passengers). If it is thought that it could be a problem then the sample size should be scaled up accordingly.

3.1.4. Example 14 – Multi-stage sampling

137. Instead of sampling every passenger on a number of selected buses, suppose we only want to sample a number of passengers on each bus. This can be thought of as multi-

stage sampling as we are sampling a number of buses (groups), and then going on to sample units (passengers) within each group.

138. We know that there are 12,000 buses and there are on average 30 passengers on each bus, of which we plan to sample 15. From a small pilot study we already know the following:

Bus	Proportion of passengers that would have travelled by bus
1	0.37
2	0.46
3	0.28
4	0.52

139. The equation for the number of buses to be sampled is:

$$c \geq \frac{\frac{SD_B^2}{\bar{p}^2} \times \frac{M}{M-1} + \frac{1}{u} \times \frac{SD_w^2}{\bar{p}^2} \times \frac{(\bar{N} - \bar{u})}{(\bar{N} - 1)}}{\frac{0.1^2}{1.96^2} + \frac{1}{M-1} \frac{SD_B^2}{\bar{p}^2}} \tag{Equation (55)}$$

Where:

- C** = Number of groups that should be sampled
- M** = Total number of groups in the population (12,000 buses)
- \bar{u}** = Number of units to be sampled within each group (pre-specified as 15 passengers)
- \bar{N}** = Average units per group (30 passengers on each bus)
- SD_B^2** = Unit variance (variance between buses)
- SD_w^2** = Average of the group variances (average within bus variation)
- \bar{p}** = Overall proportion
- 1.96** = Represents the 95% confidence required
- 0.1** = Represents the 10% absolute precision

140. Using our table of pilot information we can calculate the unknown quantities for the equation above:

Bus	Proportion of passengers that would have used a bus p_i	Variance within bus $p_i(1-p_i)$
1	0.37	0.2331
2	0.46	0.2484
3	0.28	0.2016
4	0.52	0.2496
Variance	$SD_B^2 = 0.0110$	

Bus	Proportion of passengers that would have used a bus p_i	Variance within bus $p_i(1-p_i)$
Average	$\bar{p} = 0.41$	$SD_w^2 = 0.2332$

Where:

\bar{p} is the average proportion of passengers who travel by bus, i.e. $\bar{p} = \frac{0.37 + \dots + 0.52}{4} = 0.41$

SD_w^2 is the average variance between passengers on n a bus, i.e.

$$SD_w^2 = \frac{0.2331 + \dots + 0.2496}{4} = 0.2332$$

SD_B^2 is the variance between the bus proportions, i.e. the variance between 0.37, 0.48, etc. This can be calculated in the usual way for calculating a variance, i.e. using the equation

$$SD_B^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} \text{ which gives } SD_B^2 = 0.0110$$

141. Substituting our values into the group sample size equation gives:

$$c \geq \frac{\frac{0.0110}{0.41^2} \times \frac{12000}{(12000-1)} + \frac{1}{15} \times \frac{0.2332}{0.4075^2} \times \frac{(30-1)}{(30-1)}}{\frac{0.1^2}{1.96^2} + \left(\frac{1}{(12000-1)} \times \frac{0.0110}{0.41^2} \right)} = 44.0 \tag{Equation (56)}$$

142. Therefore if we were to sample 15 passengers from each bus we should sample 44 buses for the required confidence/precision. The table below gives the number of buses required (c) when choosing to sample different numbers of passengers from each bus (u).

Number of passengers sampled on each bus u	Required number of buses c
5	119
10	63
15	44
20	35
30	26

143. The required sample size from the cluster sampling scheme example was 26; this is the same as the sample size required under the multi-stage sampling scheme when u=30 (the number of passengers sampled from each bus). This is because we assumed an

average of 30 passengers on each bus in the calculations, so when we take u =assumed average passengers the two sampling schemes are the same.

3.2. Mean value parameter of interest (Transport project)

144. This section covers sample size calculations where the objective of the project relates to a mean value of interest, under four different sampling schemes. For the sample size calculations, we need to know:

- (a) The expected mean (the desired reliability is expressed in relative terms to the mean);
- (b) The standard deviation;
- (c) The level of precision, and confidence in that precision (95/10 for all large-scale examples).

145. The parameter of interest in the examples below is average journey length (km) of people who travel by car, whether this is a domestic car or a taxi, and for people who travel by bus.

3.2.1. Example 15 – Simple random sampling

146. Suppose we are interested in the average distance (km) of car journeys in Bogota in a day (including both domestic cars and taxis), and we assume that the journeys are homogeneous. We know that 2,000,000 journeys are made each day and believe that the mean is 8 km with a standard deviation of 3.5 km. Using simple random sampling the sample size equation is:

$$n = \frac{1.96^2 NV}{(N-1) \times 0.1^2 + 1.96^2 V} \quad \text{Equation (57)}$$

Where:

$$V = \left(\frac{SD}{mean} \right)$$

n = Sample size

N = Total number of journeys (2,000,000)

$mean$ = Expected mean journey length (8 km)

SD = Expected standard deviation for the journey length (3.5 km)

1.96 = Represents the 95% confidence required

0.1 = Represents the 10% precision

$$V = \left(\frac{3.5}{8} \right)^2 = 0.19 \quad \text{Equation (58)}$$

$$n = \frac{1.96^2 \times 2,000,000 \times 0.19}{(2,000,000 - 1) \times 0.1^2 + 1.96^2 \times 0.19} = 73.5 \quad \text{Equation (59)}$$

- 147. Therefore the required sample size is at least 74 journeys to find the average journey length with 95% confidence and a 10% relative margin of error.
- 148. The calculation above does not take into account non-response. If a level of non-response is expected within the sample then the sample size should be scaled up accordingly. For example if we expected 95% of the people on journeys sampled to respond then we should take this into account and plan to sample $74/0.95 = 78$ journeys instead of 74.
- 149. There is an approximate equation for this sample size calculation. Please see the "Approximate equation" section under **ii) CFL Project – Mean value parameter of interest, Example 5: Simple random sampling** for notes relating to approximate equations. Note that 1.96 should be used in place of 1.645 to account for the increased confidence required in the large-scale projects.

3.2.2. Example 16 – Stratified random sampling

- 150. The fundamental aspect of this sampling scheme is that the average journey length differs between domestic cars and taxis (it is not homogeneous as assumed in the previous example). Because we know that the type of vehicle affects the journey distance, we want to make sure that we sample a representative number of domestic cars and taxis. A summary of each stratification group is given below:

Stratification group	Number of journeys per day	Mean (km)	Standard Deviation (km)
Domestic Car	1 595 169	9	3.7
Taxi	982 224	7	2.5

- 151. Using the data in the table above we can estimate the overall mean and standard deviation:

Overall mean:

$$mean = \frac{(g_a \times m_a) + (g_b \times m_b) + (g_c \times m_c) + \dots + (g_k \times m_k)}{N} \quad \text{Equation (60)}$$

Where:

- mean* = Weighted overall mean
- g_i* = Size of the *i*th group where *i*=1,...,k
- m_i* = Mean of the *i*th group where *i*=1,...,k
- N* = Population total

152. Substituting the values from our example into the above expression gives:

$$\begin{aligned} \text{mean} &= \frac{(1595169 \times 9) + (982224 \times 7)}{(1595169 + 982224)} && \text{Equation (61)} \\ \text{mean} &= 8.2 \end{aligned}$$

153. Overall Standard Deviation:

$$SD = \sqrt{\frac{(g_a \times SD_a^2) + (g_b \times SD_b^2) + (g_c \times SD_c^2) + \dots + (g_k \times SD_k^2)}{N}} \quad \text{Equation (62)}$$

Where:

SD = Weighted overall standard deviation

SD_i = Standard deviation of the ⁱth group where i=1,...,k, (note that these are all squared – so the group size is actually being multiplied by the group variance)

154. Using the values from our example gives:

$$\begin{aligned} SD &= \sqrt{\frac{(1595169 \times 3.7^2) + (982224 \times 2.5^2)}{(1595169 + 982224)}} && \text{Equation (63)} \\ SD &= 3.3 \end{aligned}$$

155. The sample size equation uses the overall mean and standard deviation calculated above:

$$n \geq \frac{1.96^2 \times NV}{(N-1) \times 0.1^2 + 1.96^2 V} \quad \text{Equation (64)}$$

156. Substituting in the values from our examples gives:

$$V = \left(\frac{SD}{\text{mean}} \right)^2 = \left(\frac{3.3}{8.2} \right)^2 = 0.16 \quad \text{Equation (65)}$$

$$n = \frac{1.96^2 \times 2577393 \times 0.16}{(2577393-1) \times 0.1^2 + 1.96^2 \times 0.16} = 61.4 \quad \text{Equation (66)}$$

157. This gives us the total number of journeys that should be sampled across both vehicle types. The section below assumes proportional allocation – which means that the number of journeys we want to sample from each vehicle type is proportional to the number of journeys made by each vehicle type within the population.

Equation (67)

General equation:
$$n_i = \frac{g_i}{N} \times n$$

158. Domestic cars: $n_{Car} = \frac{1592169}{2577393} \times 62 = 38.4$ Taxis: $n_{Taxi} = \frac{982224}{2577393} \times 62 = 23.6$

159. Rounding these figures results in a sample consisting of 39 domestic car journeys, and 24 taxi journeys. The summation of these group sample sizes (39 + 24 = 63) is slightly greater than that calculated from the equation above due to rounding.

160. The sample size calculated above assumes 100% response, and therefore needs to be scaled up where non-response is likely to occur.

161. This sample size is smaller than that from simple random sampling in this example. This is due to the standard deviations within strata being smaller than the standard deviation across the whole population (which is usually the case).

3.2.3. Example 17 – Cluster sampling

162. Now consider a different scenario. The parameter of interest is the average journey length of people who take local buses. Instead of sampling numerous individual passengers we would like to sample everyone from a few buses (clusters). Knowing that there are 12,000 local buses in Bogota each day, how many buses would we have to sample to find the average journey length with 95/10 confidence/precision?

163. The equation used to give us the required number of clusters, *c*, to sample is:

$$c \geq \frac{1.96^2 MV}{(M-1) \times 0.1^2 + 1.96^2 V}$$

Equation (68)

Where:

$$V = \left(\frac{SD}{\text{Cluster mean}} \right)^2$$

c = Number of clusters (buses) to be sampled

M = Total number of clusters (buses)

1.96 = Represents the 95% confidence required

0.1 = Required precision (the equation takes into account that this is relative)

164. To perform the calculations we need information about journey length at the bus level, i.e. total journey length aggregated across all passengers on a bus. If such information does not already exist, we might collect it in a pilot study. The example here assumes that data are available from four buses.

Bus	Total journey length (on average) ⁹
A	195
B	96
C	63
D	159

165. Calculating the mean and standard deviation for this total journey length for a bus gives us:

$$\text{Cluster mean (i.e. } \bar{y}) = \frac{1}{n} \sum_{i=1}^n y_i = \frac{195 + 96 + 63 + 159}{4} = 128.25 \quad \text{Equation (69)}$$

$$SD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} = \sqrt{\frac{(196-128.25)^2 + \dots + (159-128.25)^2}{3}} = 59.7180 \quad \text{Equation (70)}$$

166. These statistics (i.e. mean and SD of data) are easily produced using standard statistical software. Substituting these values into the equation gives the required number of clusters, i.e. buses as:

$$V = \left(\frac{59.7180}{128.25} \right)^2 = 0.22 \quad \text{Equation (71)}$$

$$c \geq \frac{1.96^2 \times 12000 \times 0.22}{(12000-1) \times 0.1^2 + 1.96^2 \times 0.22} = 82.7 \quad \text{Equation (72)}$$

167. The total number of buses that should be sampled is 83. Asking the journey lengths from everyone on each of the 83 buses sampled will satisfy the 95/10 confidence/precision criterion.

3.2.4. Example 18 – Multi-stage sampling

168. Continuing the previous example, suppose that we want to sample a number of local buses, but we only want to sample a number of individuals on each bus (unlike cluster sampling where everyone on each selected bus is sampled). This is an example of multi-stage sampling, as we are sampling a number of groups (buses), and we are then going on to sample a number of units (passengers) from each selected group (bus).

169. We start by assuming that we want to sample five passengers on each selected bus. In general terms we will call this number u (for units).

⁹ These totals may be derived from collecting data on all individual passengers on a bus, or otherwise by taking a sample of them and scaling up from the sample to all the passengers on the bus.

170. In order to be able to perform a sample size calculation we need information on:
- (a) The variation between individual passengers on the bus;
 - (b) The variation between buses;
 - (c) The average journey length for a passenger;
 - (d) The average journey length at the bus level (when aggregated across all passengers).
171. A previous study had provided data for passengers on three different buses, and the results are summarized below.
172. Note that not all buses will have exactly the same number of passengers.

Journey length (km)				
Bus	Number of passengers	Mean journey length ¹⁰	Total journey length aggregated over all passengers on the bus	Standard deviation ¹¹ (between passengers on the same bus)
A	26	6.9	179	3.30
B	21	7.5	157	6.21
C	30	6.7	200	3.78
Total number of passengers	77			
Overall mean journey length per passenger		7.0		
Mean total journey length (per bus)			179	
SD _B = Standard deviation between buses (SD of the total journey length column)			4 889	
SD _W = Average between passenger (within bus) standard deviation				4.44

173. In the table above, the overall mean journey length is the average length per passenger, i.e.:

$$\text{Overall mean} = \frac{179 + 157 + 200}{77} = 7.0 \quad \text{Equation (73)}$$

174. The cluster mean journey length is the mean length per bus, i.e.

¹⁰ This can be a mean from all passengers on the bus or a mean from a sample of passengers.

¹¹ And this can be a standard deviation based on all passengers or from a sample of passengers.

$$\text{Cluster mean} = \frac{179+157+200}{3} = 179 \quad \text{Equation (74)}$$

175. SD_w^2 is the average variance between passengers within buses. Its square root (SD_w) is the average within bus standard deviation. The equation for SD_w^2 is:

$$SD_w^2 = \frac{26 \times 3.30^2 + 21 \times 6.21 + 30 \times 3.78^2}{77} = 19.75 \quad \text{Equation (75)}$$

and so $SD_w = 4.44$

176. SD_B^2 is the variance between the mean journey lengths per bus. Its square root is the standard deviation between buses. It can be calculated using the general equation for a variance:

$$SD_B^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} \quad \text{where the } y_i \text{ are the journey lengths for the different buses.}$$

$$SD_B^2 = 467 \quad \text{and so } SD_B = 22 \quad \text{Equation (76)}$$

177. As well as the information from the table above we also require the number of buses, and the average number of passengers on each bus for the whole population. For this example we are using 12,000 buses with an average of 30 passengers on each bus.

$$c \geq \frac{\left(\frac{SD_B}{\text{Clustermean}} \right)^2 \times \left(\frac{M}{M-1} \right) + \left(\frac{1}{u} \right) \times \left(\frac{SD_w}{\text{Overallmean}} \right)^2 \left(\frac{\bar{N}-u}{\bar{N}-1} \right)}{\left(\frac{0.1}{1.96} \right)^2 + \frac{1}{M-1} \left(\frac{SD_B}{\text{Clustermean}} \right)^2} \quad \text{Equation (77)}$$

Where:

- M = Total number of groups (12,000 buses)
- \bar{N} = Average number of units per group (30 passengers per bus)
- u = Number of units that have been pre-specified to be sampled per group (pre-specified number of passengers to be sampled on each bus = 5)
- 1.96 = Represents the 95% confidence required
- 0.1 = Required precision

$$c \geq \frac{\left(\frac{22}{179}\right)^2 \times \left(\frac{12000}{12000-1}\right) + \left(\frac{1}{5}\right) \times \left(\frac{4.44}{7.0}\right)^2 \left(\frac{30-5}{30-1}\right)}{\left(\frac{0.1}{1.96}\right)^2 + \frac{1}{12000-1} \left(\frac{22}{179}\right)^2} = 32.6 \quad \text{Equation (78)}$$

- 178. Therefore if we were to sample five passengers from each bus we should sample 33 buses for the required confidence/precision.
- 179. Producing a table such as the one below, with different values of *u*, can help decide the practicalities of allocating limited resources, while still satisfying the 95/10 confidence/precision criterion.

Number of passengers sampled on each bus <i>u</i>	Required number of buses <i>c</i>
5	33
10	17
15	12
20	9
25	7

- 180. In this example, by doubling the number of passengers on each bus to be sampled from 5 to 10, we substantially reduce the number of buses that need to be sampled from 33 to 17.
- 181. Note that in the above example the numbers of passengers on a bus were different for the different buses. In practice this is likely to be the case, although the actual numbers may not always be known. This is not critical to the sample size calculation. What is important is that sensible estimates of the mean and standard deviation at both the cluster level (bus level) and unit level (passenger level) are used in the calculation.

3.3. Mean value parameter of interest (transport project)

- 182. This section covers an example sample size calculation based on systematic sampling where the objective of the project relates to a mean value of interest.
- 183. As for all absolute parameter of interest examples we need to know:
 - (a) The expected mean (the desired reliability is expressed in relative terms to the mean);
 - (b) The standard deviation;
 - (c) The level of precision, and confidence in that precision (95/10 for all large-scale examples).
- 184. The parameter of interest for the example below is the average journey time of buses on a specific route. We know that over a month 960 journeys are made on the route of interest, and the average journey time is 18 minutes with a standard deviation of 6 minutes.

3.3.1. Example 19 – Systematic sampling

185. Using systematic sampling we want to sample every n^{th} journey on that route. The sample size equation for a required 95/10 confidence/precision is:

$$n \geq \frac{1.96^2 V}{0.1^2} \quad \text{Equation (79)}$$

$$\text{Where: } V = \left(\frac{SD}{mean} \right)^2$$

186. Substituting in mean and standard deviation from above gives:

$$V = \left(\frac{6}{18} \right)^2 = 0.11 \quad \text{Equation (80)}$$

$$n \geq \frac{1.96^2 \times 0.11}{0.1^2} = 42.7 \quad \text{Equation (81)}$$

187. In total we should sample 43 journey times. We want to take these samples evenly spread over the 960 journeys made each month, therefore we should sample one journey for every N/n – that is one journey every 22 (= 960 / 43).
188. We can make sure that we sample at random by selecting a random starting point between 1 and 22, say journey 18, and then sample every 22nd journey from this point on: 18, 40, 62, 84, 106, etc. up to 960. This would give us a sample evenly spread over the month that is large enough to estimate the average journey time with 95/10 confidence/precision.
189. It may be more practical to sample every 20th journey rather than every 22nd. This would result in more samples being taken than the 43 calculated above – the only effect this would have would be to increase the precision and so would be perfectly acceptable.

Appendix 2. Best-practice examples for a single sampling plan for a homogeneous PoA

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

1. The “Standard for sampling and surveys for CDM project activities and programme of activities” (hereinafter called the Sampling standard) allows for a single sampling plan for two different situations, i.e. CPAs are homogeneous relative to the parameter of interest, or the CPAs are not homogeneous relative to the parameters of interest. In the latter situation, the heterogeneity can be taken into account for the analysis of the survey results. The simpler case would be that all the CPAs in the group can be shown to be homogeneous. However, the choice of a single sampling plan does not necessarily mean all the CPAs need to be homogeneous. There can be several CPAs, all following the same plan but the CPAs could be considered as different with respect to certain parameters of interest.
2. A straightforward approach that takes into account the heterogeneity between CPAs with respect to the parameters of interest, is to consider a stratified sampling approach. That is, a stratified analysis is applied systematically for all CPAs under a single sampling plan. The sampling standard and sampling guidelines contain requirements and examples for the application of stratified sampling techniques. This is because, in the presence of heterogeneity between strata, stratification yields smaller variance estimates than simple random sampling design, hence is more likely to achieve the required reliability criterion. One or more criteria from the below list of criteria that are not exhaustive may be used by the project proponents to show homogeneity to apply single sampling plan:
 - (a) Project technology/equipment have comparable input/output characteristics, including efficiency, and provide comparable service, e.g. CFLs operating on grid electricity (this would be the case when there is no significant difference between the grids involved within the boundary, whereas for certain parameters such as the retention rates of CFLs, results from stable grids may differ from the rates estimated on weak grids that suffer from voltage and frequency fluctuations);
 - (b) End users of the project technology/equipment have comparable socioeconomic conditions (e.g. middle class households);
 - (c) The geographic locations of project equipment do not have a significant influence on the parameter of interest, e.g. biogas digesters installed in colder climates have different output rates than those in warm climates, in which case stratification of digesters by geographical area is desirable;
 - (d) Installation dates of the CPAs are not significantly different to considerably impact on the parameter of interest; for example when the retention rates of CFLs are being investigated, it may be possible to show by applying equation 3 of AMS-II.J that the commissioning dates of CFLs within the CPAs chosen do not show failure rates of over 5 per cent within the time periods chosen.

2. Description of an example PoA “Improved cookstoves in country X”

3. An example PoA is the dissemination of efficient and improved cookstoves for households in country X. Three CPAs have already been registered under the PoA, which are CPA 1 “Improved cookstoves in province A”, CPA 2 “Improved cookstoves in province B” and CPA 3 “Improved cookstoves in province C”. A small-scale methodology, “AMS II.G. Energy efficiency measures in thermal applications of non-renewable biomass”, is applied to the PoA.
4. The PoA aims to reduce greenhouse gas (GHG) emissions by replacing traditional three-stone stoves with improved cookstoves. The replacement of traditional stoves by improved cookstoves increases heat transfer to the cooking utensil, thereby reducing the amount of fuel (non-renewable biomass) required for cooking. A reduction in consumption of non-renewable biomass contributes towards a reduction in GHGs emitted. Thus, improved cookstoves reduce GHG emissions by virtue of their higher thermal efficiency compared to traditional stoves.
5. Improved cookstoves were distributed in three CPAs. Relevant dates for each CPA and the monitoring period for each CPA and the entire PoA are described in the table below. Each improved cookstove installed is assigned a unique serial number for easy identification and avoidance of double counting.

CPA	Number of improved cookstoves	Date of first improved cookstove installed under the CPA/PoA	Date of last improved cookstove installed under the CPA/PoA	Monitoring period of the CPA/PoA
CPA 1	30 000	1 April 2011	31 May 2011	1 June 2011 until 31 Dec. 2012
CPA 2	10 000	1 April 2011	30 June 2011	1 July 2011 until 31 Dec. 2012
CPA 3	20 000	1 April 2011	31 July 2011	1 Aug. 2011 until 31 Dec. 2012
Total	60 000	1 April 2011	31 July 2011	1 June 2011 until 31 Dec. 2012

6. To calculate emission reductions achieved in the monitoring period of each CPA as shown in the table above), the following parameters were to be determined using ex post sampling survey:
 - (a) Percentage of improved cookstoves that are still in use (hereinafter called “retention rate of improved cookstove”);
 - (b) Percentage¹ of displaced traditional cookstoves that continue to be used (hereinafter called “continued-use rate of displaced traditional cookstove”);
 - (c) Operating efficiency of distributed improved cookstove (η_{new});
7. As stated above, according to the sampling standard, a single sampling plan covering a group of CPAs may be undertaken applying a 95/10 confidence/precision criterion for the sample size calculation. That is, the populations of all CPAs in the group are combined together, the sample size is determined and a single survey is undertaken to

¹ As per paragraph 26 (b) of the methodology “AMS-II.G: Energy efficiency measures in thermal applications of non-renewable biomass” (version 05), if traditional stoves continue to be used, monitoring shall ensure that the fuel-wood consumption of those stoves is excluded from *Bo/d*.

collect data according to whether included CPAs are assumed homogeneous relative to the concerned parameter or not. The former case requires a simple random sampling plan, the latter case requires a stratified sampling plan.

8. In the following example, the three CPAs included are assumed to be homogeneous with respect to three target parameters at the time of the survey for the following reasons:
 - (a) Only one type of improved cookstove is distributed;
 - (b) Beneficiaries of improved cookstoves are domestic households, i.e. the PoA does not target commercial users;
 - (c) Timing of the project implementation is within three months among the CPAs.
9. Since the three parameters of interest are assumed to be the same in each CPA at the time of sampling survey during the monitoring period, a single survey was undertaken using a single random sampling plan. The populations of all three CPAs are combined together (i.e. 60,000 as a total population for the survey) and then the sample size is calculated using a simple random sampling plan as shown below.

3. Sample size calculations using a simple random sampling scheme

10. According to the “Standard for sampling and surveys for CDM project activities and programme of activities”, if there is more than one parameter to be estimated, then a sample size calculation should be done for each of them. Then either the largest number for the sample size is chosen as sampling effort with one common survey, or separate sampling efforts and surveys are undertaken for each parameter.
11. Also, there are different equations to calculate a required sample size for different situations. Which equation to use depends on the type of parameter of interest, that is either:
 - (a) A percentage – in this example, retention rate of improved cookstoves and continued-use rate of displaced traditional cookstoves; or
 - (b) A numeric value – in this example, the mean value of operating efficiency of improved cookstove.
12. For all of the examples, we require 95% confidence that the margin of error in our estimate is not more than $\pm 10\%$ in relative terms, because a single sampling plan is chosen.

3.1. Retention rate of improved cookstove (proportion parameter)

13. The required sample size for this parameter is dependent on:
 - (a) The expected value of the proportion parameter;
 - (b) The level of precision and confidence in that precision (95/10 reliability criterion for a single sampling plan).
14. Based on similar studies done in the region, it is thought that this proportion is 0.85 (85%) during this monitoring period. Since the CPAs are assumed to be homogeneous

with respect to the retention rate of improved cookstove, a simple random sampling plan is conducted to estimate the proportion parameter of interest.

CPA	Number of improved cookstoves	Expected
CPA 1	30 000	
CPA 2	10 000	
CPA 3	20 000	
Total	60 000	0.85

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

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

Where:

- n = Sample size
 N = Total number of households (60,000)
 p = Expected proportion (0.85)
 1.96 = Represents the 95% confidence required
 0.1 = Represents the 10% relative precision ($0.1 \times 0.85 = 0.085 = 8.5\%$ points either side of p)

16. Substituting in our values gives:

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

17. Therefore the required sample size is at least 68.

18. Note that the figure of 68 cookstoves means 68 cookstoves with data for analysis. If we expected the response rate from the sampled cookstoves to be only 70% then we would need to scale up this number accordingly. Thus we would decide to sample $68/0.70 = 98$ cookstoves.

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

19. The required sample size for this parameter is dependent on:

- The expected value of the proportion parameter;
- The level of precision and confidence in that precision (95/10 reliability criterion for a single sampling plan).

20. Based on similar studies done in the region, it is thought that this proportion is 0.10 (10%) during this monitoring period. Since the CPAs are assumed to be homogeneous

with respect to the continued-use rate of displaced traditional cookstove, a simple random sampling plan is conducted to estimate the proportion parameter of interest.

CPA	Number of improved cookstoves	Expected
CPA 1	30 000	
CPA 2	10 000	
CPA 3	20 000	
Total	60 000	0.10

21. According to the “Standard for sampling and surveys for CDM project activities and programme of activities”, “a proportion can describe either of the two possible scenarios of the success rate or the failure rate. For example (i) cookstove still operational or (ii) cookstove no longer operational, and project proponents may use the larger of the two proportions in the sample size calculation, which is p or $(1-p)$. In this case, a 10% continued-use rate is equivalent to a 90% discontinued use. The sample size calculation is therefore based on anticipating a discontinued use of 90%.”
22. Substituting our values into equation 2 gives:

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

23. Therefore the required sample size is at least 43.
24. Note that the figure of 43 cookstoves means 43 cookstoves with data for analysis. If we expected the response rate from the sampled cookstoves to be only 70% then we need to scale up this number accordingly. Thus we would decide to sample $43/0.70 = 62$ cookstoves.

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

25. The required sample size for this parameter is dependent on:
- The expected mean;
 - The expected standard deviation;
 - The level of precision and confidence in that precision (95/10 reliability criterion for a single sampling plan).
26. Based on similar studies done in the region, it is thought that the expected mean is 0.20 with an expected standard deviation of 0.05 during this monitoring period. Since the CPAs are assumed to be homogeneous with respect to the operating efficiency of distributed improved cookstove, a simple random sampling plan is conducted to estimate the mean parameter of interest.

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

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

$$n \geq \frac{1.96^2 NV}{(N-1) \times 0.1^2 + 1.96^2 \times V} \tag{Equation (4)}$$

Where:

- V = $\left(\frac{SD}{mean}\right)^2$
- n = Sample size
- N = Total number of households
- Mean* = Expected mean (0.20)
- SD* = Expected standard deviation (0.05)
- 1.96 = Represents the 95% confidence required
- 0.1 = Represents the 10% relative precision

$$V = \left(\frac{0.05}{0.20}\right)^2 = 0.0625 \tag{Equation (5)}$$

$$n = \frac{1.96^2 \times 60,000 \times 0.0625}{(60,000 - 1) \times 0.1^2 + 1.96^2 \times 0.0625} = 24.0 \tag{Equation (6)}$$

- 28. Therefore the required sample size is at least 24 cookstoves.
- 29. Again, if we expected the response rate from the sampled cookstoves to be only 70% then we would need to scale up the number obtained above accordingly. Thus we would decide to sample $24/0.70 = 34$ cookstoves.
- 30. Oversampling is strongly encouraged at the design stage, not only to compensate for any attrition, outliers or non-response associated with the sample, but also to prevent a situation at the analysis stage where the required reliability is not achieved and additional sampling efforts would be required. According to the standard for sampling and surveys, if there is more than one parameter to be estimated in a survey, the required sample size has to be the largest number obtained in the calculations.

Therefore, one sample of 98 households to which cookstoves were distributed is selected for all three parameters of interest.

Parameter	Required minimum sample size, adjusted for the response rate
Retention rate of improved cookstove	98
Continued-use rate of displaced traditional cookstove	62
Operating efficiency of distributed improved cookstove	34

4. Sampling results and reliability calculations

31. The examples presented below illustrate how to estimate a numeric parameter and a proportion and how to quantify their reliability. The sampling method used is simple random sampling.
32. A total of 98 samples were selected to cover the combined population of CPA 1, CPA 2 and CPA 3. The adopted required reliability criteria is 95/10, i.e. 95% confidence and 10% precision.

4.1. Retention rate of improved cookstove (proportion parameter)

33. The parameter of interest is the proportion (or percentage) of improved cookstoves that were still operational in the whole population, and it is estimated from the sample that was collected. The target population are the 60,000 improved cookstoves that were distributed. A simple random sample of 98 of these improved cookstoves was taken, and for each it was recorded whether or not it was still operational. A random number generator was used to select 98 cookstoves.
34. In this example, there were 92 cookstoves out of the 98 that were still in operation in the period that was monitored. The sample proportion is therefore $p = \frac{92}{98} = 0.9388$. Rounding this to two decimal places gives us a proportion of 0.939. In other words, 93.9% of the cookstoves were still operational.
35. To assess whether or not the desired reliability was achieved, the results need to be expressed using a confidence interval. In this example the 95% confidence interval for the population proportion is 0.892 to 0.986, i.e. from 89.2% to 98.6%.

$$p \pm 1.96 \times \sqrt{1 - \frac{n}{N}} \times \sqrt{\frac{p(1-p)}{n}} = \text{Equation (7)}$$

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

$$0.939 \pm 0.047 = (0.892, 0.986)$$

Where:

n = Sample size

- p = Sample proportion
 1.96 = Represents the 95% confidence required

36. We are therefore 95% confident that the range from 89.3% to 98.7% contains the true population percentage of improved cookstoves that are still operational in this time period.
37. Half the width of the confidence interval (also known as margin of error) can now be used to assess whether or not the selected reliability criterion has been met. Put simply, reliability is achieved if the distance from the upper confidence limit to the sample proportion (or equally if the distance from the sample proportion to the lower confidence limit) is within 10% of the sample proportion in relative terms.
38. Reliability can therefore be checked using the following calculation:

$$\frac{\frac{1}{2}\text{width of confidence interval}}{\text{estimated proportion}} \times 100\% \quad \text{Equation (8)}$$

39. Here the proportion of cookstoves that are still operational is 0.939, with a 95% confidence interval of 0.892 to 0.986, i.e. from 89.2% to 98.6%. The achieved reliability is therefore:

$$\frac{0.986 - 0.939}{0.939} \times 100\% = \frac{0.939 - 0.892}{0.939} \times 100\% = \frac{0.047}{0.939} \times 100\% = 5.0\% \quad \text{Equation (9)}$$

40. The achieved precision is 5% of the sample proportion and so the sample estimate of 93.9% operational cookstoves meets the required specification.

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

41. The parameter of interest is the proportion (or percentage) of traditional cookstoves that were no longer used in the target population, and it is estimated by the sample proportion. The target population are the 60,000 cookstoves. A simple random sample of 98 of these cookstoves was taken, and for each it was recorded whether or not the traditional cookstove was still being used. The same 98 households were surveyed as for the estimation of the previous parameter.
42. In this example, there were 88 households out of the 98 where the traditional cookstoves were no longer used. The sample proportion is therefore $p = \frac{88}{98} = 0.898$. In other words, 89.8% of the traditional cookstoves were no longer used.
43. Again the results of sampling need to be summarized using a confidence interval. In this example, the 95% confidence interval for the proportion is 0.838 to 0.958, i.e. from 83.8% to 95.8%.

$$p \pm 1.96 \times \sqrt{1 - \frac{n}{N}} \times \sqrt{\frac{p(1-p)}{n}} = \text{Equation (10)}$$

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

$$0.898 \pm 0.059 = (0.838, 0.958)$$

Where:

- n = Sample size
 p = Sample proportion of traditional cookstoves no longer being used
 1.96 = Represents the 95% confidence required

44. We are therefore 95% confident that the range from 83.8% to 95.8% contains the true population percentage of cookstoves that are no longer used in this time period.

45. As above, reliability can be checked using the following calculation:

$$\frac{\frac{1}{2} \text{width of confidence interval}}{\text{proportion}} \times 100\% \quad \text{Equation (11)}$$

46. Here the proportion of traditional cookstoves that are no longer used is 0.898, with a 95% confidence interval of 0.838 to 0.958. Reliability is therefore:

$$\frac{0.958 - 0.898}{0.898} \times 100\% = \frac{0.898 - 0.838}{0.898} \times 100\% = 6.7\% \quad \text{Equation (12)}$$

47. The achieved precision is 6.7% of the sample proportion and so the sample estimate of 89.8% traditional cookstoves no longer used meets the selected reliability criterion.

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

48. The parameter of interest is the mean operating efficiency for the population of improved cookstoves that were distributed, and it is estimated from the sample proportion. The target population are the 60,000 cookstoves that were distributed. A simple random sample of 98 cookstoves was taken, and the operating efficiency was recorded. The common sample of 98 cookstoves used for estimating the two previous parameters was surveyed.

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

50. As above, the results need to be summarized using a confidence interval. In this example, the 95% confidence interval is 0.185 to 0.209.

$$\begin{aligned} \text{Sample Mean} \pm t_{(n-1)} \times \sqrt{1 - \frac{n}{N}} \times \frac{SD}{\sqrt{n}} &= && \text{Equation (13)} \\ 0.1971 \pm 1.985 \times \sqrt{1 - \frac{98}{60,000}} \times \frac{0.06}{\sqrt{98}} &= \\ 0.1971 \pm 0.012 &= (0.185, 0.209) \end{aligned}$$

Where:

n	=	Sample size
Mean	=	Sample mean
SD	=	Sample standard deviation
$t_{(n-1)}$	=	Relevant multiplier from the t-distribution with (n-1) degrees of freedom that is associated with 95% confidence The degrees of freedom here are 97 and the corresponding t-value is 1.985

51. We are 95% confident that the range from 0.185 to 0.209 units contains the true population mean operating efficiency of operational improved cookstoves.
52. A symmetric confidence interval for the true population parameter can also be expressed as sample mean \pm margin of error. The estimate of 0.1971 is regarded as reliable if its margin of error is within the pre-specified reliability precision (i.e. 10% of the sample mean).
53. Reliability can therefore be checked using the following calculation:

$$\frac{\frac{1}{2} \text{width of confidence interval}}{\text{mean}} \times 100\% \quad \text{Equation (14)}$$

54. Here the mean operating efficiency is 0.1971 units, and the 95% confidence interval is 0.185 to 0.209. Reliability is therefore:

$$\frac{0.209 - 0.1971}{0.1971} \times 100\% = \frac{0.1971 - 0.1850}{0.1971} \times 100\% = \frac{0.0120}{0.1971} \times 100\% = 6.1\% \quad \text{Equation (15)}$$

55. The precision is 6.1% of the sample mean and so the sample estimate of 0.1971 is within the required specification.

5. Conclusions

56. Sample estimates of all three parameters are within the required reliability precision.

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

1. Introduction

1. When the cookstoves are rolled out in three different provinces, there could be a strong possibility that the parameter of interest is not homogeneous across the three CPAs. The adoption of new technology may vary from community to community for several reasons such as: different demographics, different traditions, or the amount of input in a province (in this example, three times as many cookstoves were distributed in CPA 1 compared to CPA 2).
2. When the assumption of homogeneity across the three CPAs cannot be assumed, then a stratified analysis can be used to account for these differences in the analysis stage to manage the precision. Similar to the simple random sampling example, a single survey can be conducted for the three parameters of interest (namely retention rate, continued-use rate and operating efficiency) across the three CPAs. The difference with stratification is that the number of households sampled within each CPA is specified in advance rather than being the results of a simple random sampling process. The sample size for each CPA has been calculated for each parameter of interest below.

2. Sample size calculations

3. According to the “Standard for sampling and surveys for CDM project activities and programme of activities”, if there is more than one parameter to be estimated, then a sample size calculation should be done for each parameter. Then, either the largest number for the sample size is chosen for the sampling effort with one common survey, or the sampling effort and survey is repeated for each parameter.
4. Also, there are different equations to calculate a required sample size for different situations. Which equation to use depends on the type of parameter of interest, that is either:
 - (a) A proportion – in this example, retention rate of improved cookstoves and continued-use rate of displaced traditional cookstoves; or
 - (b) A numeric value – in this example, the mean value of operating efficiency of improved cookstove.
5. For all examples, we require 95% confidence that the margin of error is no more than $\pm 10\%$ of the estimated parameter.

2.1. Retention rate of improved cookstove (proportion parameter)

6. The required sample size for this parameter depends on:
 - (a) The value that the proportion is expected to take;

- (b) The level of precision, and confidence in that precision (95:10 confidence:precision for a single sampling plan).
7. To illustrate stratified random sampling, the estimated retention proportions for each CPA have been generated to reflect an overall retention proportion of approximately 0.85 while being heterogeneous between CPAs.

CPA	Number of improved cookstoves	Expected
CPA 1	30 000	0.84
CPA 2	10 000	0.73
CPA 3	20 000	0.91
Total	60 000	0.845

8. The formula for the total sample size is:

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

Where:

$$V = \frac{SD^2}{\bar{p}^2} = \frac{\text{Weighted overall expected variance}}{\text{Weighted overall expected proportion, squared}} \quad \text{Equation (2)}$$

We use the figures from the table above to calculate the overall variance¹ and proportion of cookstoves still in use to be used in equation 2:

$$SD^2 = \frac{(g_a \times p_a(1-p_a)) + (g_b \times p_b(1-p_b)) + (g_c \times p_c(1-p_c)) + \dots + (g_k \times p_k(1-p_k))}{N} \quad \text{Equation (3)}$$

$$\bar{p} = \frac{(g_a \times p_a) + (g_b \times p_b) + (g_c \times p_c) + \dots + (g_k \times p_k)}{N} \quad \text{Equation (4)}$$

Where:

- g_i = Size of the i^{th} group (CPA) where $i=a, \dots, k$
- p_i = Proportion of the i^{th} group (CPA) where $i=a, \dots, k$
- N = Population total

9. Substituting our values into equations 3 and 4 gives:

¹ The variance of a proportion is calculated as: $p(1-p)$.

$$SD^2 = \frac{(30,000 \times 0.84(1 - 0.84)) + (10,000 \times 0.73(1 - 0.73)) + (20,000 \times 0.91(1 - 0.91))}{60,000} = 0.12735 \quad \text{Equation (5)}$$

$$\bar{p} = \frac{(30,000 \times 0.84) + (10,000 \times 0.73) + (20,000 \times 0.91)}{60,000} = 0.845 \quad \text{Equation (6)}$$

Substituting the results from equations 5 and 6 into equation 2 gives:

$$V = \frac{0.12735}{0.845^2} = 0.1784 \quad \text{Equation (7)}$$

We can now substitute V and N into equation 1:

$$n \geq \frac{1.96^2 \times 60,000 \times 0.1784}{(60,000 - 1) \times 0.1^2 + 1.96^2 \times 0.1784} = 68.44 \quad \text{Equation (8)}$$

10. Therefore the required sample size across all three CPAs is 69.

11. We can now use the formula below to calculate the required sample size for each CPA:

$$n_i = \frac{g_i}{N} \times n \quad \text{Equation (9)}$$

Where:

- n_i = The sample size required for the i^{th} group (CPA) where $i=a, \dots, k$
- g_i = Size of the i^{th} group (CPA) where $i=a, \dots, k$
- N = Population total
- n = The total sample size required

12. Substituting the group, total and sample size into equation 9 for each CPA gives:

CPA	Number of improved cookstoves	Equation	Sample size required
CPA 1	30 000	$n_{CPA1} = \frac{30,000}{60,000} \times 69 = 34.50$	35
CPA 2	10 000	$n_{CPA2} = \frac{10,000}{60,000} \times 69 = 11.50$	12
CPA 3	20 000	$n_{CPA3} = \frac{20,000}{60,000} \times 69 = 23.00$	23

CPA	Number of improved cookstoves	Equation	Sample size required
Total	60 000		70

Note that the total sample size has increased from 69 to 70 due to the rounding up of the sample sizes within each CPA.

13. Note that the sample size of 70 requires the analysis of 70 cookstoves with valid data. If the expected response rate from the sampled cookstoves is 70%, then we need to scale up the sample size accordingly within each CPA, which is done by dividing the sample size in each CPA by the expected response rate. Note that this method allows for different response rates within each CPA. The required sample sizes adjusted for a response rate of 70% in each CPA are given below:

CPA	Sample size required (not accounting for non-response)	Equation	Sample size required, adjusted for non-response
CPA 1	35	$\frac{35}{0.7} = 50.00$	50
CPA 2	12	$\frac{12}{0.7} = 17.14$	18
CPA 3	23	$\frac{23}{0.7} = 32.86$	33
Total	70		101

14. Having adjusted for non-response within each CPA, the total sample size required for 95:10 confidence:precision for the estimate of the retention rate is 101 cookstoves.

2.2. Continued-use rate of displaced traditional cookstove (proportion parameter)

15. The required sample size for this parameter is dependent on:
- The value that the proportion is expected to take;
 - The level of precision, and confidence in that precision (95:10 confidence:precision for a single sampling plan).
16. Based on studies conducted in the region, it is thought that the overall continued-use proportion is 0.1 (10%) during the monitoring period of interest. The stratified sampling example below uses continued-use proportions generated for each CPA to give an overall proportion of about 10% while being heterogeneous between CPAs.

CPA	Number of improved cookstoves	Expected Continued-use proportions
CPA 1	30 000	0.11
CPA 2	10 000	0.18
CPA 3	20 000	0.06

CPA	Number of improved cookstoves	Expected Continued-use proportions
Total	60 000	0.105

17. In this example, the overall 10% continued-use proportion is equivalent to a 90% discontinued-use proportion, and project proponents may use the larger of the two proportions in the sample size calculations. Therefore we base the calculations on the discontinued-use proportions for each CPA below:

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

18. The equation to give us the required total sample size across all three CPAs is:

$$n \geq \frac{1.96^2 NV}{(N-1) \times 0.1^2 + 1.96^2 V} \tag{Equation (10)}$$

Where:

$$V = \frac{SD^2}{\bar{p}^2} = \frac{\text{Weighted overall expected variance}}{\text{Weighted overall expected proportion, squared}} \tag{Equation (11)}$$

We use the figures from the table above to calculate the overall variance² and proportion of cookstoves still in use to be used in equation 11:

$$SD^2 = \frac{(g_a \times p_a(1-p_a)) + (g_b \times p_b(1-p_b)) + (g_c \times p_c(1-p_c)) + \dots + (g_k \times p_k(1-p_k))}{N} \tag{Equation (12)}$$

$$\bar{p} = \frac{(g_a \times p_a) + (g_b \times p_b) + (g_c \times p_c) + \dots + (g_k \times p_k)}{N} \tag{Equation (13)}$$

Where:

- g_i = Size of the i^{th} group (CPA) where $i=a, \dots, k$
- p_i = Proportion of the i^{th} group (CPA) where $i=a, \dots, k$
- N = Population total

² The variance of a proportion is calculated as: $p(1-p)$.

19. Substituting our values into equations 12 and 13 gives:

$$SD^2 = \frac{(30,000 \times 0.89(1 - 0.89)) + (10,000 \times 0.82(1 - 0.82)) + (20,000 \times 0.94(1 - 0.94))}{60,000} = 0.09235 \quad \text{Equation (14)}$$

$$\bar{p} = \frac{(30,000 \times 0.89) + (10,000 \times 0.82) + (20,000 \times 0.94)}{60,000} = 0.8950 \quad \text{Equation (15)}$$

Substituting the results from equations 14 and 15 into equation 11 gives:

$$V = \frac{0.9235}{0.8950^2} = 0.1153 \quad \text{Equation (16)}$$

We can now substitute V and N into equation 10:

$$n \geq \frac{1.96^2 \times 60,000 \times 0.1153}{(60,000 - 1) \times 0.1^2 + 1.96^2 \times 0.1153} = 44.26 \quad \text{Equation (17)}$$

20. Therefore the required sample size is 45.

21. Having calculated the total sample size, we now calculate the required sample sizes within each CPA using the following equation:

$$n_i = \frac{g_i}{N} \times n \quad \text{Equation (18)}$$

Where:

- n_i = The sample size required for the i^{th} group (CPA) where $i=a, \dots, k$
- g_i = Size of the i^{th} group (CPA) where $i=a, \dots, k$
- N = Population total
- n = The total sample size required

22. Substituting the group, total and sample sizes into equation 18 for each CPA gives:

CPA	Number of improved cookstoves	Equation	Sample size required
CPA 1	30 000	$n_{CPA1} = \frac{30,000}{60,000} \times 45 = 22.50$	23
CPA 2	10 000	$n_{CPA2} = \frac{10,000}{60,000} \times 45 = 7.50$	8

CPA	Number of improved cookstoves	Equation	Sample size required
CPA 3	20 000	$n_{CPA3} = \frac{20,000}{60,000} \times 45 = 15.00$	15
Total	60 000		46

Note: The total sample size has increased from 45 to 46 due to the rounding up of the sample sizes within each CPA.

23. Again the sample size of 46 cookstoves may need to be adjusted upwards to account for non-responses. If a response rate of 70% is expected then the sample size within each CPA should be increased by dividing the sample size within the CPA by the expected response rate:

CPA	Sample size required (not accounting for non-response)	Equation	Sample size required adjusted for non-response
CPA 1	23	$\frac{23}{0.7} = 32.86$	33
CPA 2	8	$\frac{8}{0.7} = 11.43$	12
CPA 3	15	$\frac{15}{0.7} = 21.43$	22
Total	46		67

24. Having adjusted for non-response within each CPA, the total sample size required for 95:10 confidence:precision for the estimate of the continued-use rate is 67 cookstoves.

2.3. Operating efficiency of distributed improved cookstove (mean value parameter)

25. The required sample size for this parameter is dependent on:
- The expected mean, since the desired reliability is expressed in relative terms to the mean;
 - The expected standard deviation;
 - The level of precision, and confidence in that precision (95:10 confidence:precision for a single sampling plan).

26. Similar studies in the region have estimated the mean operating efficiency of the cookstoves to be 0.2, with a standard deviation of 0.05. For this stratified sampling example the means and standard deviations for each CPA have been generated to reflect an overall mean and standard deviation of approximately 0.2 and 0.05 respectively while being heterogeneous across CPAs. These values are given below:

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

27. The equation to give us the required total sample size across all three CPAs is:

$$n \geq \frac{1.96^2 NV}{(N-1) \times 0.1^2 + 1.96^2 V} \quad \text{Equation (19)}$$

Where:

$$V = \frac{SD^2}{\bar{m}^2} = \frac{\text{weighted overall expected variance}}{\text{weighted overall expected mean, squared}} \quad \text{Equation (20)}$$

We use the figures from the table above to calculate the overall mean and variance:

$$\bar{m} = \frac{(g_a \times m_a) + (g_b \times m_b) + (g_c \times m_c) + \dots + (g_k \times m_k)}{N} \quad \text{Equation (21)}$$

$$SD^2 = \frac{(g_a \times SD_a^2) + (g_b \times SD_b^2) + (g_c \times SD_c^2) + \dots + (g_k \times SD_k^2)}{N} \quad \text{Equation (22)}$$

Where:

g_i Size of the i^{th} group (CPA) where $i=a, \dots, k$

m_i Mean of the i^{th} group (CPA) where $i=a, \dots, k$

SD_i Standard deviation of the i^{th} group (CPA) where $i=a, \dots, k$, (note that these are all squared – so the group size is actually being multiplied by the group variance)

N Population total

28. Substituting our values into equations 21 and 22 gives:

$$\bar{m} = \frac{(30,000 \times 0.20) + (10,000 \times 0.05) + (20,000 \times 0.30)}{60,000} = 0.2083 \quad \text{Equation (23)}$$

$$SD^2 = \sqrt{\frac{(30,000 \times 0.06^2) + (10,000 \times 0.09^2) + (20,000 \times 0.03^2)}{60,000}} = 0.00345 \quad \text{Equation (24)}$$

Substituting the results from equations 23 and 24 into equation 20 gives:

$$V = \frac{0.00345}{0.2083^2} = 0.0795 \quad \text{Equation (25)}$$

We can now substitute V and N into equation 19:

$$n \geq \frac{1.96^2 \times 60,000 \times 0.0795}{(60,000 - 1) \times 0.1^2 + 1.96^2 \times 0.0795} = 30.5211 \quad \text{Equation (26)}$$

29. Therefore the required total sample size is 31 cookstoves.

30. We can use the formula below to calculate the size of the sample for each CPA:

$$n_i = \frac{g_i}{N} \times n \quad \text{Equation (27)}$$

Where:

- n_i = The sample size required for the i^{th} group (CPA) where $i=a, \dots, k$
- g_i = Size of the i^{th} group (CPA) where $i=a, \dots, k$
- N = Population total
- n = The total sample size required

31. Substituting the CPA and total populations, and the calculated total sample size into equation 27 for each CPA gives:

CPA	Number of improved cookstoves	Equation	Sample size required
CPA 1	30 000	$n_{CPA1} = \frac{30,000}{60,000} \times 31 = 15.50$	16

CPA	Number of improved cookstoves	Equation	Sample size required
CPA 2	10 000	$n_{CPA2} = \frac{10,000}{60,000} \times 31 = 5.17$	6
CPA 3	20 000	$n_{CPA3} = \frac{20,000}{60,000} \times 31 = 10.33$	11
Total	60 000	31	33

Note: The total sample size has increased from 31 to 33 due to the rounding up of the sample sizes within each CPA.

32. Finally, if the expected response rate is 70%, the sample sizes must be increased by dividing the sample size in each group by 0.7, to ensure that the estimate of the operating efficiency will have the required level of precision. The required sample sizes adjusted for a response rate of 70% are given below for each CPA:

CPA	Sample size required (not accounting for non-response)	Equation	Sample size required adjusted for non-response
CPA 1	16	$\frac{16}{0.7} = 22.86$	23
CPA 2	6	$\frac{6}{0.7} = 8.57$	9
CPA 3	11	$\frac{11}{0.7} = 15.71$	16
Total	33		48

33. Having adjusted for non-response within each CPA the total sample size required for 95:10 confidence:precision for the estimate of the mean cookstove operating efficiency is 48.

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

When estimating more than one parameter in a single survey, the sample size used for the survey should be the maximum sample size calculated across all of the parameters. This ensures that the required confidence:precision levels are met for all parameters. Considering the sample size calculations for the three previous stratified sampling plan examples, the sample size for a single survey which will collect data on all three parameters should be 101 cookstoves.

Parameter	Required minimum sample size, adjusted for a 70% response rate
Retention rate of improved cookstove	101
Discontinued-use rate of displaced traditional cookstove	67
Operating efficiency of distributed improved cookstove	48

Note: The breakdown of the total sample size of 101 cookstoves into the three CPAs is as shown in the retention proportion example.

34. A sample of 101 cookstoves stratified across the three CPAs will produce combined population estimates of the three parameters at the 95% confidence level with at least 10% precision. For those parameters that require a smaller sample size, the effect of this over-sampling will be an increase in the precision associated with the estimates of these parameters. Therefore, using these examples, sampling 101 cookstoves will result in greater than 10% precision for the estimates of the discontinued-use and operating efficiency estimates.

3.1. Retention rate of improved cookstove (proportion parameter)

35. For this example, a stratified sample of 101 cookstoves across three CPAs were surveyed. The parameter of interest is the proportion of improved cookstoves still in operation in the population of 60,000 improved cookstoves that were initially distributed. The table below details the number of improved cookstoves to be sampled within each CPA along with the number that were actually sampled, because of attrition and non-response.

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

36. The table below shows the simulated survey results:

CPA	Number of cookstoves surveyed	Number of cookstoves still in operation	Retention proportion
CPA 1	42	35	$p_{CPA1} = \frac{35}{42} = 0.83$
CPA 2	12	9	$p_{CPA2} = \frac{9}{12} = 0.75$
CPA 3	25	23	$p_{CPA3} = \frac{23}{25} = 0.92$

37. Here, the overall proportion of cookstoves still in operation across all three CPAs needs calculating, rather than the individual CPA proportions. The overall proportion is a weighted average of the proportions within each CPA and is calculated using the equation below:

$$p_{Strat} = \sum_{i=a}^k \frac{g_i}{N} \times p_i \quad \text{Equation (28)}$$

Where:

p_{Strat}	=	The stratified estimated overall proportion
g_i	=	Size of the i^{th} group (CPA) where $i=a, \dots, k$
N	=	Population total
p_i	=	Proportion observed in the i^{th} group (CPA) where $i=a, \dots, k$

Substituting in the observed proportions in each group and population totals gives:

$$p_{Strat} = \left(\frac{30,000}{60,000} \times 0.83 \right) + \left(\frac{10,000}{60,000} \times 0.75 \right) + \left(\frac{20,000}{60,000} \times 0.92 \right) = 0.8467 \quad \text{Equation (29)}$$

38. Therefore the estimated overall proportion of improved cookstoves still in operation across the three CPAs is 0.8467, or 84.67% if expressed as a percentage.
39. To quantify the level of precision associated with this estimate, a confidence interval needs to be calculated. As sample size calculations were based on a 95% confidence level, we calculate the confidence interval using the equation below:

$$p_{Strat} \pm 1.96 \times s.e. (p_{Strat}) \quad \text{Equation (30)}$$

Where:

p_{Strat}	=	The stratified estimated overall proportion
1.96	=	Z value corresponding to 95% confidence level
$s.e. (p_{Strat})$	=	Standard error of the stratified estimated overall proportion

We can calculate the standard error for the stratified estimated overall proportion with:

$$s.e. (p_{Strat}) = \sqrt{\sum_{i=a}^k \left(\frac{g_i}{N} \right)^2 \times \left(1 - \frac{n_i}{g_i} \right) \times \frac{p_i(1 - p_i)}{n_i}} \quad \text{Equation (31)}$$

Where:

$s.e. (p_{Strat})$	=	Standard error of the stratified estimated overall proportion
g_i	=	Size of the i^{th} group (CPA) where $i=a, \dots, k$
N	=	Population total
n_i	=	Number of sampled units in the i^{th} group (CPA) where $i=a, \dots, k$

p_i Proportion observed in the i^{th} group (CPA) where $i=a, \dots, k$

Substituting in the values for our example we can calculate the standard error as follows:

$$s.e.(p_{Strat}) = \sqrt{\left(\left(\frac{30,000}{60,000}\right)^2 \times \left(1 - \frac{42}{30,000}\right) \times \frac{0.83(1-0.83)}{42}\right) + \left(\left(\frac{10,000}{60,000}\right)^2 \times \left(1 - \frac{12}{10,000}\right) \times \frac{0.75(1-0.75)}{12}\right) + \left(\left(\frac{20,000}{60,000}\right)^2 \times \left(1 - \frac{25}{30,000}\right) \times \frac{0.92(1-0.92)}{25}\right)}$$

Equation (32)

$$s.e.(p_{Strat}) = 0.0398$$

40. Substituting the results from equations 29 and 32 into equation 30 we find that the 95% confidence interval is:

$$0.8467 \pm 1.96 \times 0.0398 = (0.7687, 0.9247)$$

Equation (33)

41. Therefore the proportion of improved cookstoves in operation across the three CPAs is estimated to be 0.8467, with a 95% confidence interval of 0.7687 to 0.9247.

42. Expressed in relative terms, reliability is quantified using the following equation:

$$Reliability = \frac{0.5 \times Width\ of\ Confidence\ Interval}{Estimated\ Stratified\ Overall\ Proportion} \times 100$$

Equation (34)

Substituting the values from equation 33 into equation 34 gives:

$$Reliability = \frac{0.8467 - 0.7687}{0.8467} \times 100 = \frac{0.078}{0.8467} \times 100 = 9.21\%$$

Equation (35)

43. The achieved precision is 9.21% of the sample proportion, and so the sample estimate of 84.67% operational cookstoves meets the required specification.

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

44. Although the previous sample size calculations for this parameter of interest specified that a sample size of 67 would meet the required 95:10 confidence:precision levels, 101 cookstoves were to be sampled so that the estimate of the parameter (retention rate) met the required levels of precision. The table below details the number of cookstoves for which data is available, which was fewer than the planned 101 due to non-response:

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

45. Recall from the sample size part of this example that the sample size calculations were based on the discontinued-use proportions rather than continued-use proportions. Therefore the estimates calculated below relate to the discontinued-use proportions.
46. The table below shows the simulated number of traditional cookstoves still in use, and the discontinued-use proportion, within each CPA:

CPA	Number of cookstoves actually surveyed	Number of traditional cookstoves still in use	Number of traditional cookstoves no longer in use	Discontinued-use Proportion
CPA 1	42	6	36	$p_{CPA1} = \frac{36}{42} = 0.86$
CPA 2	12	2	10	$p_{CPA2} = \frac{10}{12} = 0.83$
CPA 3	25	1	24	$p_{CPA3} = \frac{24}{25} = 0.96$

47. Though the table above shows the estimated discontinued-use proportions for each CPA, our focus is on the overall discontinued-use proportion across all three CPAs, which can be calculated using the equation below:

$$p_{Strat} = \sum_{i=a}^k \frac{g_i}{N} \times p_i \tag{Equation (36)}$$

Where:

- p_{Strat} = The stratified estimated proportion
- g_i = Size of the i^{th} group (CPA) where $i=a, \dots, k$
- N = Population total
- p_i = Proportion observed in the i^{th} group (CPA) where $i=a, \dots, k$

Substituting in the observed proportions in each group and population totals gives:

$$p_{Strat} = \left(\frac{30,000}{60,000} \times 0.86 \right) + \left(\frac{10,000}{60,000} \times 0.83 \right) + \left(\frac{20,000}{60,000} \times 0.96 \right) = 0.8883 \tag{Equation (37)}$$

48. The estimated discontinued-use proportion across the three CPAs is 0.8883, or 88.83% if expressed as a percentage.
49. Once again we can assess the level of precision associated with this estimate in equation 37 using a confidence interval. The equation for a 95% confidence interval for this estimate is:

$$p_{Strat} \pm 1.96 \times s.e. (p_{Strat}) \tag{Equation (38)}$$

Where:

- p_{Strat} = The stratified estimated overall proportion

- 1.96 = Z value corresponding to 95% confidence level
- $s.e.(p_{Strat})$ = Standard error of the stratified estimated overall proportion

We can calculate the standard error for the stratified estimated overall proportion with:

$$s.e.(p_{Strat}) = \sqrt{\sum_{i=a}^k \left(\frac{g_i}{N}\right)^2 \times \left(1 - \frac{n_i}{g_i}\right) \times \frac{p_i(1 - p_i)}{n_i}} \quad \text{Equation (39)}$$

Where:

- $s.e.(p_{Strat})$ = Standard error of the stratified estimated overall proportion
- g_i = Size of the i^{th} group (CPA) where $i=a, \dots, k$
- N = Population total
- n_i = Number of sampled units in the i^{th} group (CPA) where $i=a, \dots, k$
- p_i = Proportion observed in the i^{th} group (CPA) where $i=a, \dots, k$

Substituting in the values for our example we can calculate the standard error as follows:

$$s.e.(p_{Strat}) = \sqrt{\left(\left(\frac{30,000}{60,000}\right)^2 \times \left(1 - \frac{42}{30,000}\right) \times \frac{0.86(1 - 0.86)}{42}\right) + \left(\left(\frac{10,000}{60,000}\right)^2 \times \left(1 - \frac{12}{10,000}\right) \times \frac{0.83(1 - 0.83)}{12}\right) + \left(\left(\frac{20,000}{60,000}\right)^2 \times \left(1 - \frac{25}{30,000}\right) \times \frac{0.96(1 - 0.96)}{25}\right)} \quad \text{Equation (40)}$$

$$s.e.(p_{Strat}) = 0.0349$$

- 50. Substituting the results from equations 37 and 40 into equation 38 the 95% confidence interval for the estimated discontinued-use proportion is:

$$0.8883 \pm 1.96 \times 0.0349 = (0.8199, 0.9567) \quad \text{Equation (41)}$$

- 51. Therefore the discontinued-use proportion across the three CPAs is estimated to be 0.8883, with a 95% confidence interval of 0.8199 to 0.9567.

- 52. Reliability, expressed as a percentage, can be assessed using the following equation:

$$Reliability = \frac{0.5 \times \text{Width of Confidence Interval}}{\text{Estimated Stratified Proportion}} \times 100 \quad \text{Equation (42)}$$

Substituting the values from equation 41 into equation 42 gives:

$$Reliability = \frac{0.8883 - 0.8199}{0.8883} \times 100 = \frac{0.0684}{0.8883} \times 100 = 7.7\% \quad \text{Equation (43)}$$

- 53. The achieved precision is 7.7% of the sample proportion, and so the sample estimate of 88.83% discontinued use of traditional cookstoves meets the required specification.

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

54. The third parameter of interest is the mean operating efficiency of the distributed improved cookstoves. As one study was conducted to collect data for all three parameters of interest, the planned and achieved sample sizes shown in the table below are the same as for the two previous examples:

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

55. The table below shows the survey results for mean operating efficiency and standard deviations for each of the CPAs:

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

56. To obtain the sample estimate of the overall mean operating efficiency across all three CPAs we need to calculate the weighted average. This can be done using equation 44:

$$m_{strat} = \sum_{i=a}^k \frac{g_i}{N} \times m_i \tag{Equation (44)}$$

Where:

- m_{Strat} = The stratified estimated overall mean
- g_i = Size of the i^{th} group (CPA) where $i=a, \dots, k$
- N = Population total
- m_i = Mean of the i^{th} group (CPA) where $i=a, \dots, k$

Substituting in the observed means in each group and population totals gives:

$$m_{strat} = \left(\frac{30,000}{60,000} \times 0.22 \right) + \left(\frac{10,000}{60,000} \times 0.06 \right) + \left(\frac{20,000}{60,000} \times 0.28 \right) = 0.2133 \tag{Equation (45)}$$

57. We can assess the level of precision associated with the estimate in equation 45 using a confidence interval. The equation for a 95% confidence interval for this estimate is:

$$m_{Strat} \pm 1.96 \times s.e. (m_{Strat}) \quad \text{Equation (46)}$$

Where:

m_{Strat}	=	The stratified estimated overall mean
1.96	=	Z value corresponding to 95% confidence level
$s.e. (m_{Strat})$	=	Standard error of the stratified estimated overall mean

We can calculate the standard error for the stratified estimated overall mean as follows:

$$s.e. (m_{Strat}) = \sqrt{\sum_{i=a}^k \left(\frac{g_i}{N}\right)^2 \times \left(1 - \frac{n_i}{g_i}\right) \times \frac{SD_i^2}{n_i}} \quad \text{Equation (47)}$$

Where:

$s.e. (m_{Strat})$	=	Standard error of the stratified estimated overall mean
g_i	=	Size of the i^{th} group (CPA) where $i=a, \dots, k$
N	=	Population total
n_i	=	Number of sampled units in the i^{th} group (CPA) where $i=a, \dots, k$
SD_i^2	=	Variance of the i^{th} group (CPA) where $i=a, \dots, k$

Substituting the values from our example into equation 47 we obtain:

$$s.e. (m_{Strat}) = \sqrt{\left(\left(\frac{30,000}{60,000}\right)^2 \times \left(1 - \frac{42}{30,000}\right) \times \frac{(0.07)^2}{42}\right) + \left(\left(\frac{10,000}{60,000}\right)^2 \times \left(1 - \frac{12}{10,000}\right) \times \frac{(0.06)^2}{12}\right) + \left(\left(\frac{20,000}{60,000}\right)^2 \times \left(1 - \frac{25}{30,000}\right) \times \frac{(0.04)^2}{25}\right)} \quad \text{Equation (48)}$$

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

58. Substituting results from equations 45 and 48 into equation 46, a 95% confidence interval for the population mean operating efficiency is:

$$0.2133 \pm 1.96 \times 0.00667 = (0.2003, 0.2264) \quad \text{Equation (49)}$$

59. The operating efficiency mean across the three CPAs is estimated to be 0.2133, with a 95% confidence interval of 0.2003, to 0.2264.

60. Reliability, expressed as a percentage, is quantified using the following equation:

$$\text{Reliability} = \frac{0.5 \times \text{Width of Confidence Interval}}{\text{Estimated Stratified Overall Mean}} \times 100 \quad \text{Equation (50)}$$

Substituting the values from equation 49 into equation 50 gives:

$$Reliability = \frac{0.2133 - 0.2003}{0.2133} \times 100 = \frac{0.0130}{0.2133} \times 100 = 6.1\% \quad \text{Equation (51)}$$

61. The achieved precision is 6.1% of the sample mean, so the sample estimate of 0.2133 mean operating efficiency of improved cookstoves meets the required specification.

4. Conclusion

In summary, sample estimates of all three parameters are within the required reliability precision.

Appendix 4. Best-practice examples for reliability calculations

1. Introductory notes on reliability calculations

1. The two examples presented here illustrate how to estimate a numeric parameter and a proportion and how to check their reliability. The sampling method used in both cases is simple random sampling. Both examples are assumed to be small-scale project activities where the required reliability criteria is 90/10, i.e. 90% confidence and 10% precision.
2. If calculations are being performed manually, it is important to retain as many decimal places as relevant, until the final calculated figure is reached. Rounding can then be carried out. To emphasize this, the calculations presented here use figures to several decimal places.

2. Example 1: CFL Project – Numeric parameter

3. The parameter of interest in this example is the mean average daily usage of a CFL (in hours) for a whole population of CFLs that were distributed in a particular region of a country.
4. The population is the 420,000 households to which CFLs were distributed, one per household. A simple random sample of 140 households was taken, and the average daily usage (in hours) of each CFL was recorded. These are presented in the table below.

Table 1 Average CFL usage (in hours)

cfl	usage	cfl	usage	cfl	usage	cfl	usage	cfl	usage	cfl	usage	cfl	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

5. The parameter of interest – the mean average daily usage of a CFL (in hours) for this whole population of CFLs – is estimated from the sample mean. This is often written as \bar{y} (and is equal to $\frac{1}{n}(y_1 + y_2 + \dots + y_n)$, or the shorthand form $\frac{1}{n} \sum_{i=1}^n y_i$). n is the sample size, i.e. 140.
6. The mean average usage for the sample of 140 CFLs is 3.4686 hours. As a simple summary this is rounded to 1 or 2 decimal places, i.e. the mean average usage of the CFLs is estimated to be 3.47 hours.

3. Confidence, precision and reliability

7. Instead of presenting just a single estimate, it is better to summarize the results of sampling using a confidence interval. In this example the 90% confidence interval is 3.22 to 3.71 hours. We are 90% sure that the true population mean value for average usage of a CFL is between 3.22 hours and 3.71 hours. Whilst the sample mean is the estimate that will be used in calculations, it is always advisable when presenting it in a report to do so along with its confidence interval.
8. The 90% confidence interval for the population mean is given by the equation: sample mean \pm t-value \times standard error of the mean.
9. The estimate of 3.47 hours is regarded as reliable if the precision of the study – as defined by the t-value \times standard error of the mean – is within the pre-specified reliability precision. For small-scale mechanisms this is 10% of the mean.
10. Detailed calculations are presented below. In this example the precision is 7.1% of the mean and so the sample estimate of 3.47 hours is within the required specification.

4. Checking reliability

4.1. Standard error of the mean

11. The equation for the standard error of the mean when data have been collected using

simple random sampling is $\sqrt{(1-f) \frac{s^2}{n}}$.

f is the sampling fraction – the proportion of the population that is sampled.

Here it is $\frac{140}{420000} = 0.00033$.

s^2 is the sample variance (s is the sample standard deviation).

For this sample of 140 CFLs, $s^2 = 3.0826$ and $s = 1.7557$.

n is the sample size, i.e. 140.

12. Putting all these pieces of information together gives:

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

Equation (1)

and so the standard error of the mean is 0.1484.

4.2. t-value

13. This value depends on (i) the level of confidence and (ii) the size of the sample. The exact figure can be acquired from statistical tables for the t-distribution, or using standard statistical software. The value can also be derived in Microsoft Excel using the TINV¹ function.

14. For a sample size of 140 the t-value is 1.6559.

4.3. Precision

15. The precision associated with an estimate is: t-value × standard error of the mean.
16. The precision of the mean average CFL usage (in hours), assuming 90% confidence, in this example is therefore: ± (1.6559 × 0.1484) i.e. ± 0.2457.
17. The ratio of this relative to the mean CFL usage is $\frac{0.2457}{3.4686} = 0.0708$ and so the relative precision is 7.1%. The data are therefore within the required specification.

4.4. Another way of checking reliability

18. The limits of the confidence interval are sample mean ± t-value × standard error of the mean, which can be written more generally as sample mean ± precision, where the lower limit is mean minus precision and the upper limit is mean + precision.
19. Reliability can therefore be checked using the following calculation:

$$\frac{\frac{1}{2}\text{width of confidence interval}}{\text{mean}} \times 100\% \quad \text{Equation (2)}$$

20. For example, here the mean CFL usage is 3.4686, and the 90% confidence interval is 3.2230 to 3.7143 hours. Reliability is therefore:

$$\frac{\frac{1}{2}(3.7143-3.2230)}{3.4686} \times 100\% = \frac{\frac{1}{2} \times 0.4913}{3.4686} \times 100\% = 7.1\% \quad \text{Equation (3)}$$

¹ TINV(0.10,(sample size minus 1)) will give the t-value associated with 90% confidence. For example here TINV(0.10,139) gives the t-value for a sample size of 140 and 90% confidence.

21. The above approach is likely to be most useful when the data have been analysed using statistical software which produced the relevant confidence interval as well as the sample mean.

5. Example 2 : Cookstove project – Proportional parameter

22. The parameter of interest in this example is the proportion (or percentage) of cookstoves in a particular region of a country that were still operational at the end of the third year after the stoves were distributed.

23. The population of interest is the 640,000 households, and there was one cookstove per household. A simple random sample of 274 of these households was taken, and for each of them it was recorded whether or not the cookstove was still operational.

24. The parameter of interest – the proportion (or percentage) of cookstoves that were still operational in the whole population – is estimated from the sample proportion.

25. This is often written as p and is calculated as $p = \frac{r}{n}$ where r is the number of “successes”, in this case the number of cookstoves that are still in operation, and n is the total number of cookstoves that are observed in the sample.

26. In this example there were 159 cookstoves out of the 274 that were still in operation. The sample proportion is therefore $p = \frac{159}{274} = 0.5803$. Rounding this to two decimal places gives us a proportion of 0.58. In other words, 58% of the cookstoves were still operational after the third year.

5.1. Confidence, precision and reliability

27. Instead of presenting just a single estimate, it is better to summarize the results of sampling using a confidence interval. In this example the 90% confidence interval for the proportion is 0.5313 to 0.6293. We are therefore 90% sure that the percentage of cookstoves in the population that are still operational is between 53% and 63%.

28. The 90% confidence interval for the population proportion is given by the equation: sample proportion $\pm 1.6449 \times$ standard error of the proportion.²

29. The estimate of 58% is regarded as reliable if the precision of the study – as defined by $1.6449 \times$ standard error of the proportion – is within the pre-specified reliability precision. For small-scale mechanisms this is 10% of the proportion. In this case ± 0.058 in absolute terms or $\pm 5.8\%$.

30. Detailed calculations are presented below. In this example the precision is 8.5% of the sample proportion and so the sample estimate of 58% operational cookstoves is within the required specification.

² A confidence interval for a proportion is: sample proportion \pm z-value \times standard error of the proportion. The z-value depends on the level of confidence. For 90% confidence it is 1.6449.

5.2. Checking reliability

5.2.1. Standard error of the proportion

31. The equation for the standard error of the proportion when data have been collected using simple random sampling is $\sqrt{(1-f)\frac{pq}{n}}$

32. f is the sampling fraction – the proportion of the population that is sampled.

33. Here it is $\frac{274}{640000} = 0.00043$

34. p is the sample proportion, i.e. 0.5803. $q = (1-p)$. It represents the proportion of cookstoves that are not operational after three years, and is 0.4197. n is the sample size, i.e. 274.

35. Putting all these pieces of information together gives:

$$\sqrt{(1-f)\frac{pq}{n}} = \sqrt{(1-0.00043)\frac{0.5803 \times 0.4197}{274}} = \sqrt{0.00089} = 0.0298 \quad \text{Equation (4)}$$

36. Note that this standard error could also be calculated using the actual numbers of population size, sample size, number of operational cookstoves etc., i.e.:

$$\sqrt{(1-f)\frac{pq}{n}} = \sqrt{\left(\frac{640000 - 274}{640000}\right)\left(\frac{159}{274}\right)\left(\frac{115}{274}\right)} = \sqrt{0.00089} = 0.0298 \quad \text{Equation (5)}$$

37. The standard error of the proportion is 0.0298. In terms of the standard error of the percentage it is 2.98%.

5.2.2. Precision

38. The precision associated with a proportion is: z-value \times standard error of the proportion. The precision of the proportion of operational cookstoves in this example, assuming 90% confidence, is: $\pm (1.6449 \times 0.0298)$ i.e. ± 0.0490 .

39. The ratio of this relative to the proportion of cookstoves that are still operational is $\frac{0.0490}{0.5803} = 0.0845$ and so the relative precision is 8.5%. The data are within the required specification.

5.2.3. Another way of checking reliability

40. The limits of the confidence interval are sample proportion \pm z-value \times standard error of the proportion, which can be written more generally as sample proportion \pm precision, where the lower limit is the proportion minus precision and the upper limit is the proportion plus precision.

41. Reliability can therefore be checked using the following calculation:

$$\frac{\frac{1}{2}\text{width of confidence interval}}{\text{proportion}} \times 100\% \quad \text{Equation (6)}$$

42. For example, here the proportion of cookstoves that are still operational is 0.5803, with a 90% confidence interval of 0.5313 to 0.6293.

43. Reliability is therefore:

$$\frac{\frac{1}{2}(0.6293 - 0.5313)}{0.5803} \times 100\% = \frac{\frac{1}{2} \times 0.0980}{0.5803} \times 100 = 8.5\% \quad \text{Equation (7)}$$

44. The above approach is likely to be most useful when the data have been analysed using statistical software which produced the relevant confidence interval and sample proportion.

5.2.4. Comments

45. The equation above assumes that the distribution of the proportion is approximately Normal. That is usually an acceptable assumption provided the proportion of interest is not too small and not too large, and the sample size is not too small;

46. If the sampling fraction f is small then the multiplier $(1 - f)$ in the above calculation will be very close to 1. In some instances, therefore, the equation that is used for the standard error of the proportion is the conservative equation $\sqrt{\frac{pq}{n}}$;³

47. If statistical software is used to undertake the calculation, the software may use the exact equation for calculating the confidence interval (which assumes a Binomial distribution as opposed to the Normal approximation). In this case the reliability would be checked using the equation which is based on the width of the confidence interval.

6. How to deal with failure to achieve reliability

6.1. Introductory notes on how to deal with failure to achieve reliability

48. This section proposes some steps to follow when the required reliability is not met by sample data. It uses as its scenario a small-scale CDM project activity where the reliability criteria is 90:10 (i.e. 90% confidence and 10% relative precision), and where the parameter of interest is a numeric one.

³ It is conservative because $\sqrt{\frac{pq}{n}}$ will be greater than $\sqrt{(1-f)\frac{pq}{n}}$.

6.2. Scenario : CFL project – Numeric parameter

49. The parameter of interest in this example is the mean average daily usage of a CFL (in hours) for a whole population of CFLs that were distributed in a particular region of a country. The population is the 420,000 households to which CFLs were distributed, one per household. A simple random sample was to be taken.
50. The sample size calculation used a value of 3.5 hours for the expected population mean CFL usage and a value of 2.5 hours for the expected population standard deviation (SD). This gave a required sample size of 138 households.

6.3. Example 1

51. A simple random sample of 140 households was taken, and the average daily usage (in hours) of each CFL was recorded. The summaries of the data and of the reliability calculations are presented below.

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	± 0.5294
Relative precision	± 0.1422 i.e. 14.22%

52. Here the required reliability is not met by the sample data and so we conclude that the mean of 3.72 hours is not sufficiently reliable. It also means that the confidence interval associated with the parameter is wider than required.
53. The 90% confidence interval is 3.19 to 4.25 hours, which is telling us that the population mean CFL usage is likely (90% likely) to be somewhere in the range of 3 hours 11 minutes to 4 hours 15 minutes.
54. Possible steps to take when addressing the problem of lack of reliability in sample data include the following, although the first one (scrutinizing data) should always be carried out before doing any calculations, in order to ensure that the data are of the highest quality and that the planned analyses are appropriate.
55. The list below offers some order to the approaches. First of all – if it has not been done already – comes scrutinizing the data; then there are some possible statistical analysis approaches which could be used on the existing data. Failing that an additional sample could be taken. There is no obvious ordering to the analysis approaches.
1. Scrutinize the raw data;
 2. Possible analysis approaches;
 - (a) Scrutinize the summary statistics;

(b) Post-stratification;

3. Take an additional sample.

56. If none of these are successful then the reliability has not been met.⁴

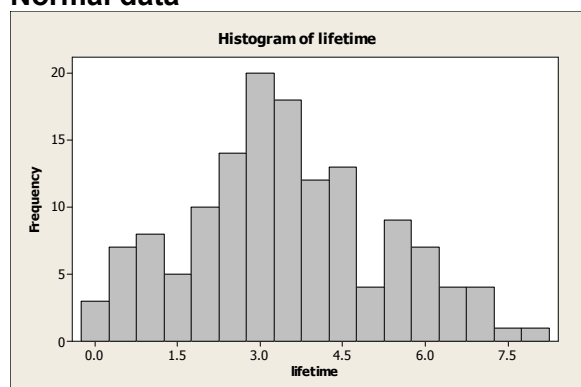
6.4. Analysis approach 1: Scrutinize the raw data

57. It is vitally important to scrutinize the raw data carefully prior to estimating the mean and checking its reliability, and this can be done using graphical summaries such as histograms, boxplots, and normal probability plots. These plots would show up outliers in the data or any skewness in the distribution of the data.

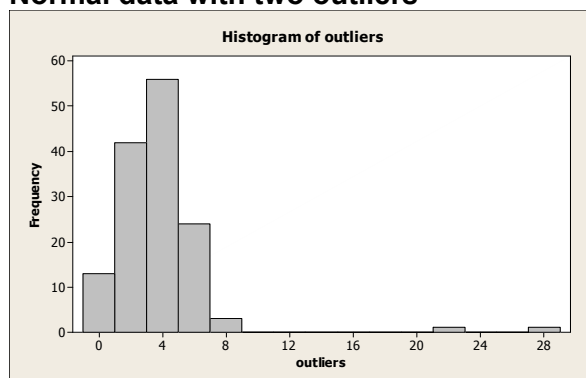
58. An **outlier** can be the result of a mistake (wrongly recorded, or wrongly entered onto the computer in which case it can be corrected); or it could be real value - in which case it must be left as it is and included in the analysis. If data are highly **skewed**, then it may be that they should be transformed prior to the analysis. The reliability would then be determined using the analysis of the transformed data. Example transformations include the logarithm, or the square root.

59. Below shows examples of a histogram for: (i) normal data (ii) data with one outlier and (iii) skewed data.

Normal data

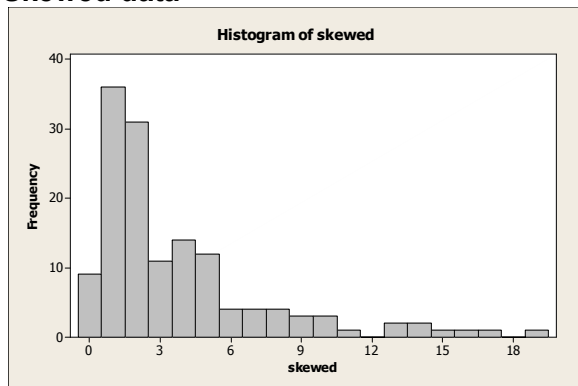


Normal data with two outliers



⁴ A generalized method of discounting to account for deficiencies in reliability is not included in this document, however where the project proponents can demonstrate that discounting of emission reduction estimates or taking the lower bound or upper bound of estimates of the parameter are the only recourse available to the project proponents, procedures for request for deviation shall be followed.

Skewed data



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

60. The two key elements to the sample size calculation (other than the reliability and confidence) are the expected population mean value and the expected population standard deviation. If the sample data fails to meet the required reliability then this could be due to the sample estimates being quite different 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 reliability will be greater than 10%;
- (b) If the sample mean is the same as the expected population mean but the standard deviation is larger than expected then the reliability will be greater than 10%.

61. This is demonstrated in the table below, where the sample size calculation used is the same as described in the Scenario on page 1 (mean CFL usage of 3.5 hours, SD of 2.5 hours giving a required sample size of 138 households).

	Population size (N)	Sample size (n)	Sample mean	Sample SD	Absolute precision	Relative precision
(i)	420 000	140	3.0 hours	2.5	0.35	11.68%
(ii)	420 000	140	3.5 hours	3.0	0.42	11.99%

62. In (i) the absolute precision is as planned but, because the sample mean is lower than expected, the relative precision is larger than the required 10%. In (ii) because the SD is larger than expected both the absolute and relative precision are larger than required by the sample size specification.

63. It seems not unreasonable to try to accept data that come from scenarios of type (i). The following rule is therefore proposed.

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

- 64. In Example 1, the sample size is over 100 but the standard deviation is considerably larger than was used in the sample size calculation (3.78 as opposed to 2.5; the sample standard deviation is more than 50% more than the one used in the sample size calculation). The sample data cannot be accepted.
- 65. Rationale behind the conclusions in the above paragraphs is explained below:
- 66. The distribution of the sample variance s^2 is as follows:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

From probability tables for the Chi-squared distribution we can see that, for samples of size 100, one would expect 95% of sample variances (s^2) to be less than $1.24 \times$ the population variance (i.e. less than $1.24 \sigma^2$). One could then argue that, provided it was not more than 24% more than the expected population variance, the sample variance can be regarded as an acceptable estimate of the population variance.

- 67. In other words, provided that the sample standard deviation was not more than 12% more than the population standard deviation then the sample standard deviation can be regarded as an acceptable estimate of the population standard deviation.
- 68. The table 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 more variability in the estimates of the population variance and consequently the wider the distribution.

Sample size and ratio of sample SD to population SD	
Sample size	95% of ratios less than this value
30	1.21
40	1.18
50	1.16
60	1.15
70	1.14
80	1.13
90	1.12
100	1.12
125	1.10
150	1.09
175	1.09
200	1.08
250	1.07
300	1.07

69. Balancing the requirement for a reasonable amount of information with a realisation that the sample standard deviation is unlikely to ever be exactly the same as the population one, then it seems not unreasonable to have a rule of thumb which states that:
- Provided the sample is large enough; and
 - Provided the sample standard deviation is only a fraction larger than the population one;

then the sample data could be accepted. It is therefore proposed that a minimum sample size be 100 units and maximum “overage” be 10%.

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

70. If there appears to be some characteristic of the population that is responsible for the apparent increase in the variability of the data – e.g. households in the urban areas are all using their CFLs for longer than the households in rural areas – then this characteristic can be regarded as a stratification variable, and the mean CFL usage recalculated using post-stratification techniques. In this example geographical area (rural or urban) is the stratification variable. Other examples might include the different times of the year when the data are collected, if usage varies according to time of year.
71. Taking the stratification variable into account should improve the precision of the estimate of the mean usage for the whole population, and hence its reliability. It might also give a more accurate estimate of the population mean usage. An example of post-stratification is explained below.
72. This uses the same example as Example 1 where the following data were collected.

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

73. The total population of 420,000 in this part of the country are distributed across the rural and urban areas in a ratio of 60% to 40% respectively; and it is thought that CFL usage is lower in the rural areas than in the urban. Of the 140 households in the sample, it transpires that 104 of them were from rural areas and 36 from urban areas. The proportions in the sample from the rural and urban areas at 74% and 26% respectively do not match those of the population, and so the estimate of average usage over all 140 CFLs may be slightly lower than it should be for the whole population.
74. Summary statistics for the two sub-groups are as follows:

Summary statistics	Rural	Urban
Population size	252 000	168 000

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

75. We can use post-stratification to estimate the mean average CFL usage to reflect the rural:urban proportions in the population by using a weighted average of the stratum means as follows:

76. Post-stratification mean $\bar{y}_{\text{post_st}} = \frac{1}{N} \sum_{h=1}^L N_h \bar{y}_h$ where N is the total population size; N_h is the population size in each stratum; \bar{y}_h is the mean for each stratum; L is the number of strata.

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

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

$$se = \sqrt{\sum_{h=1}^L \left(\frac{N_h}{N} \right)^2 \left(\frac{N_h - n_h}{N_h} \right) \frac{s_h^2}{n_h}}$$

where N, N_h are as above; and s_h is the stratum standard deviation.

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

and equals 0.2616.

78. Consequently the absolute precision (assuming a 90% confidence interval) is $1.645 \times 0.2616 = 0.4304$ and so the relative precision is 0.4304 divided by the mean of 4.7024 i.e. 9.15%. The reliability has been met.

6.7. Analysis approach 3: Take an additional sample

79. Another option to improve the precision of the study data is to take an additional sample. The formula for the size of this additional sample would be the same as for the actual study. However, the value(s) that should be used for the standard deviation (SD) and the mean should be different. It is also advisable to try different combinations of the values (sample and population ones) in order to identify a total sample size which will be large enough to address the reliability concern. The additional sample size will then be the difference between this figure and the originally planned sample size.

80. 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 CFLs gives a mean of 3.72 hours and a standard deviation of 3.78. These figures could therefore be

more realistic estimates of either the mean daily hours of CFL usage or the standard deviation, and so can be used in new sample size calculations. The table below presents a few different combinations of mean and standard deviation and their resulting sample sizes. The first line in the table (in italics) is the original calculation.

Sample size for 90:10 reliability (90% confidence and 10% precision)		
SD	Mean	Sample size
<i>2.5</i>	<i>3.5</i>	<i>138</i>
2.5	3.72	123
3.78	3.5	316
3.78	3.72	280

- 81. If the project proponent (PP) believes that the population mean really is about 3.5 hours, but the standard deviation was underestimated originally and is likely to be closer to 3.78 then the required total sample size for the study should have been 316. An additional sample of $316 - 140 = 176$ is therefore required. However, if they think that the mean was also too low originally and it is likely to be about 3.72 hours then the additional sample size would be $280 - 140 = 140$.
- 82. Note that the above illustration only uses two figures for the mean and two for the standard deviation. In practice a range of different possibly relevant means and standard deviations should be used. This is no different to the practice recommended when the sample size calculation is being performed in the first place!

6.8. Alternative post-adjustment methods to deal with missed reliability targets

- 83. Alternatively, the PP may want to take the upper or lower bound of the 90% confidence interval.

6.8.1. Scenario : CFL project – numeric parameter

- 84. This uses the same example as Example 1 where the following data were collected:

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

- 85. Here the required reliability is not met by the sample data and so we conclude that the mean of 3.72 hours is not sufficiently reliable. It also means that the confidence interval associated with the parameter is wider than required.

86. In this example the 90% confidence interval is:

$$Mean \pm t_{(n-1)} \times \frac{SD}{\sqrt{n}} = 3.7230 \pm 1.6559 \times \frac{3.7838}{\sqrt{140}} \quad \text{Equation (8)}$$

Where:

N = Sample size

Mean = Sample mean

SD = Sample standard deviation

t_(n-1) = Represents the relevant value from the t-distribution with (n-1) degrees of freedom that is associated with 90% confidence⁵
The degrees of freedom here are 139 and the corresponding t-value is 1.6559

87. We are 90% sure that the true population mean value for CFL usage is between 3.19 and 4.25. Whilst the sample mean is the estimate that will be used in calculations, it is always advisable when presenting it in a report to do so along with its confidence interval.

88. Instead of taking additional samples, the PP may take the upper or lower bound of the 90% confidence interval. The choice of the upper or lower bounds to be used in estimating emission reductions shall be conducted in a manner that ensures conservativeness.

89. The lower and upper confidence limit for 90% confidence interval are calculated below:

	Lower confidence limit	Mean	Upper confidence limit
90% confidence interval	3.19	3.72	4.25

90. In order to ensure conservativeness in estimation of emission reduction, 3.19 is selected.

6.8.2. Scenario: cookstove project – proportional parameter

91. This is another example where the parameter of interest is a proportional one.

92. The parameter of interest here is the percentage of improved cookstoves that are still in use (hereinafter called the “retention rate of improved cookstove”) for a whole population of cookstoves that were distributed in a particular region of a country. The population is the 80,000 households to which cookstoves were distributed, one per household. A simple random sample was to be taken.

93. The sample size calculation used a value of 0.85 (85%) for the expected proportion. This gave a required sample size of 68 cookstoves. Considering that the expected response rate from the sampled cookstoves would be 70%, it was decided to scale up this number and to sample $68/0.70 = 98$ cookstoves. A simple random sample of 98 households was

⁵ If a large sample has been taken then this t-value can be replaced by the figure of 1.645 (for a confidence level is 90%).

taken, and the number of improved cookstoves that are still in use was recorded. The summaries of the data and of the reliability calculations are presented below:

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

94. Here the required reliability is not met by the sample data and so it is concluded that the retention rate of improved cookstoves is not sufficiently reliable. In this case study the 95% confidence interval is:

$$p \pm 1.96 \times \sqrt{\frac{p(1-p)}{n}} = 0.7551 \pm 1.96 \times \sqrt{\frac{0.7551(1-0.7551)}{98}} \quad \text{Equation (9)}$$

Where:

- N = Sample size
- P = Sample proportion
- 1.96 = Represents the 95% confidence required

95. It is 95% certain that the percentage of cookstoves that are still operational in this period is between 67.00% and 84.02%. While the sample mean is the estimate that will be used in calculations, it is always advisable when presenting it in a report to do so along with its confidence interval.

96. Instead of taking additional samples, the PP may take the upper or lower bound of the 95% confidence interval. The choice of the upper or lower bounds to be used in estimating emission reductions shall be conducted in a manner that ensures conservativeness. The lower and upper confidence limits for 95% confidence interval are calculated below:

	Lower confidence limit	Mean	Upper confidence limit
95% confidence interval	67.00%	75.51%	84.02%

97. In order to ensure conservativeness in estimation of emission reduction, 67.00% is selected.

Appendix 5. Best-practice examples – acceptance sampling

1. Introductory notes on acceptance sampling

1. The aim of this section is to demonstrate the process of selecting a validation/verification sample and using it to decide whether or not the PP's data are valid. The methodology is based on a procedure commonly known as Acceptance Sampling. It is illustrated using a survey example where the aim is to estimate a numeric parameter.

2. Example: CFL project – numeric parameter

2. The parameter of interest is the mean average daily usage of a CFL (in hours) for the whole population of CFLs that were distributed in a particular region of a country. The population is the 420,000 households to which CFLs were distributed, one per household. A simple random sample of 140 households was taken by the PP, and the average daily usage (in hours) of each CFL was recorded. From these data the PP determines a mean average daily usage.
3. To validate/verify the PP's sample the DOE needs to take, and observe, a simple random sample of households from the PPs sample. The decision about whether or not the PPs data are valid will depend on the number of discrepancies there are between the DOE's data and the PP's data.¹ The DOE therefore needs to set up, using their own professional judgement, criteria for deciding what constitutes a discrepancy. The DOE also needs to decide – again using their own professional judgement – the following:
 - (a) The proportion of discrepancies between the PP's data and DOE's data that can be considered acceptable in their sample. This is referred to as the AQL (Acceptable Quality Level);
 - (b) The proportion of discrepancies between the PP's data and DOE's data that would be considered unacceptable in their sample. This is the UQL (Unacceptable Quality Level).
4. In this example we will assume that the AQL is 1% and the UQL is 10%, though these could be different for different types of study.
5. The process of determining the size of the DOE's sample also requires what are referred to as the producer's risk and the consumer's risk,² both of which are set at 5% according to the current sampling standard.³

¹ If the data for a household in the DOE's sample is different from the data for that same household in the PP's sample that constitutes a discrepancy.

² Producer's risk is the chance that the DOE will wrongly reject the PP's dataset (i.e. reject a dataset of acceptable quality). Consumer's risk is the chance that the DOE will wrongly accept the PP's dataset i.e. accept a dataset which is unacceptable as defined above in (b).

³ The calculations require a value of acceptable quality associated with the producer's risk and a value of unacceptable quality associated with the consumer's risk. The values between the AQL and UQL need to be seen in this context.

6. All this information then determines:
 - (a) The size of the DOE's sample; and
 - (b) The acceptance number.
7. The acceptance number is the number of acceptable discrepancies. The DOE needs to observe no more than this number of discrepant records for the PP's data to be valid.
8. The calculation can be carried out using reliable statistical software, published tables of acceptance sampling standards, or hand calculations. Table 1 below may not be complete for every study that requires validation/verification, but it provides the required sample size and acceptance number for several different scenarios.

Table 1. Sample size and acceptance number⁴

AQL	UQL	Sample Size	Acceptance number
1%	10%	61	2
1%	15%	30	1
1%	20%	22	1
0.5%	10%	46	1
0.5%	15%	30	1
0.5%	20%	22	1

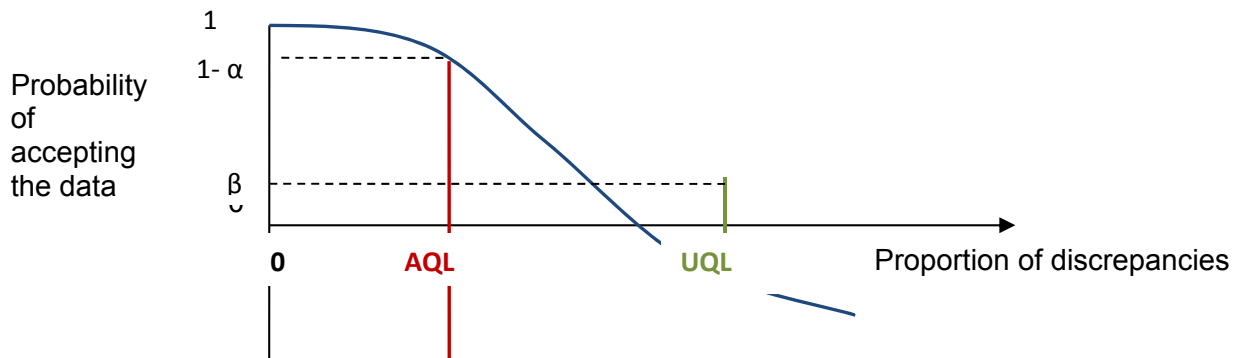
9. In this case the size of the DOE's sample is 61 households, and the acceptance number is 2. Hence, if it transpires that there are more than two records in the DOE's sample that do not agree with the PP's then the PP's data are not accepted. If there are none, one or two discrepant records then the PPs data are valid.

2.1. Worked example

10. The calculation can be seen in terms of the operating curve – i.e. the probability of accepting the PPs data for different proportions of discrepancies between their data and the DOEs – where we want:
 - (a) A high chance of accepting the PP's data when it is of acceptable quality, i.e. Probability (Proportion of discrepancies with DOE's data is less than the AQL) $\geq 1 - \alpha$;
 - (b) A low chance of accepting the PP's data when it is of unacceptable quality, i.e. Probability (Proportion of discrepancies with the DOE's is more than the UQL) $\geq \beta$.
11. Note that the *producer's risk* is the opposite of (i); i.e. it is the chance of not accepting the PP's data when it is of acceptable quality, and is equal to α . The *consumer's risk* is (ii), i.e. β .

⁴ The table is based on both the Producer's risk and the Consumer's risk being 5%.

12. In this example the AQL is 1% (or 0.01 in proportion terms) with a producer's risk $\alpha = 0.05$; and the UQL is 10% (or 0.1 in proportion terms) with a consumer's risk $\beta = 0.05$.



13. We need to find a sample size and acceptance number that satisfy (or nearly satisfy) these probability statements. Owing to the discreteness of the data it may not be possible to satisfy them exactly.
14. The approach is based on the Chi-square distribution with $2(c+1)$ degrees of freedom. It determines the acceptance number first of all and then the sample size.

2.2. Step 1: Acceptance number

15. Let

$$r(c) = \frac{\chi_{1-\beta}^2}{\chi_{\alpha}^2} \tag{Equation (10)}$$

Where: χ_{α}^2 is the 100α percentile and $\chi_{1-\beta}^2$ the $100(1-\beta)$ percentile of the χ^2 distribution with $2(c+1)$ degrees of freedom.

Then c is the smallest value satisfying:

$$r(c-1) > \frac{UQL}{AQL} > r(c) \tag{Equation (11)}$$

16. Here the ratio of UQL/AQL is 10 and so we need to find a value of c that satisfies the above.
17. Since acceptance numbers are going to be small we can construct a table of Chi-square values and the corresponding ratios for different values of c from $c = 0, 1, 2, 3$, etc. as in table 2 below.

Table 2. Chi-square values and the corresponding ratios for different values of c

Tabulated values of a Chi-square distribution with 2(c+1) degrees of freedom where $\alpha=0.05$ and $\beta=0.05$			
c	α	1-β	ratio
0	0.1026	5.9915	58.40
1	0.7107	9.4877	13.35
2	1.6354	12.5916	7.70
3	2.7326	15.5073	5.67
4	3.9403	18.3070	4.65
5	5.2260	21.0261	4.02

18. $c = 2$ is the smallest value which satisfies the above.

2.3. Step 2: Sample size

19. The required sample size, n , is such that:

$$\frac{\chi_{1-\beta}^2}{2 \times UQL} \leq n \leq \frac{\chi_{\alpha}^2}{2 \times AQL} \tag{Equation (12)}$$

Where: χ_{α}^2 and $\chi_{1-\beta}^2$ are defined as before, but now $c = 2$, and the χ^2 distribution has $2(c+1) = 6$ degrees of freedom.

20. With $c = 2$ this is:

$$\frac{12.59159}{2 \times 0.1} \leq n \leq \frac{1.635383}{2 \times 0.01} \tag{Equation (13)}$$

i.e. $62.96 \leq n \leq 81.77$.

21. So we have a sample size of 63 and an acceptance number of 2.

2.4. Step 3: Refining the calculation

22. The above steps used a Chi-square approximation, but the data actually have a Binomial distribution. The calculations can now be refined to see if the value of n could be modified. Table 3 shows the exact values of α and β , for an acceptance number of 2, and different sample sizes around 63. While the above calculation showed that we needed a sample size of 63, the table shows that sample sizes of 62 and 61 would both also have α and β below 0.05.

23. The required sample size is therefore 61 with an acceptance number of 2.

Table 3. Different sample sizes with values of α and β for an acceptance number of 2

Acceptance number (c)	Sample size (n)	Exact probabilities based on Binomial distribution	
		alpha	beta
2	60	0.022	0.053
2	61	0.023	0.049
2	62	0.024	0.045
2	63	0.025	0.042
2	64	0.027	0.039

2.5. Excel functions

24. The Excel function used in Table 2 is CHIINV(PROB, DF). It returns the value X, for a Chi-square distribution with DF degrees of freedom, where the probability of being greater than X is PROB.
25. The Excel function used in Table 3 is BINOMDIST(acceptance number, sample size, PROB, TRUE) where PROB is either the AQL or the UQL.

Appendix 6. Sample size calculator

1. The project proponents may use the sample size calculator available in the CDM website <<https://cdm.unfccc.int/Reference/Guidclarif/index.html>>.

Document information

<i>Version</i>	<i>Date</i>	<i>Description</i>
04.0	16 October 2015	EB 86, Annex 4. Revision to include recommended practices for conducting survey and data collection.
03.1	13 August 2015	Revision to include Appendix 6: sample size calculator worksheets as requested by the Board in CDM EB 80, paragraph 34.
03.0	4 October 2013	EB 75, Annex 8. Revision to include additional best practices examples for single sampling plan, pragmatic approaches to reliability calculations and a flow chart to select the most appropriate sampling plan for a given situation.
02.0	13 September 2012	EB 69, Annex 5. Revisions are: <ul style="list-style-type: none"> • To include examples to illustrate acceptance sampling as one means of verifying the sampling records of the project proponent; • To include examples for possible methods to deal with failure to achieve reliability including scrutiny of the data, post stratification, adding more sample units; • To delete examples of how to deal with very small or very large proportions in the context of simple random sampling due to the new proposed requirement on the minimum sample size; • To change the title of this draft from "Best practice examples focusing on sample size and reliability calculations and sampling for validation/verification" to "Guidelines for sampling and surveys for CDM project activities and programme of activities".
01.1	16 May 2012	Editorially revised to remove "draft" from paragraph 3 and amended title for readability.
01.0	11 May 2012	EB 67, Annex 6. Initial adoption.

Decision Class: Regulatory
 Document Type: Guideline
 Business Function: Methodology
 Keywords: DOE, calculations, programme of activities, project activities, sampling, SSC project activities, validation
