Designation: D5791 – 23

Standard Guide for Using Probability Sampling Methods in Studies of Indoor Air Quality in Buildings[1]

This standard is issued under the fixed designation D5791; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This guide covers criteria for determining when probability sampling methods should be used to select locations for placement of environmental monitoring equipment in a building or to select a sample of building occupants for questionnaire administration for a study of indoor air quality. Some of the basic probability sampling methods that are applicable for these types of studies are introduced.

1.2 Probability sampling refers to statistical sampling methods that select units for observation with known probabilities (including probabilities equal to one for a census) so that statistically defensible inferences are supported from the sample to the entire population of units that had a positive probability of being selected into the sample.

1.3 This guide describes those situations in which probability sampling methods are needed for a scientific study of the indoor air quality in a building. For those situations for which probability sampling methods are recommended, guidance is provided on how to implement probability sampling methods, including obstacles that may arise. Examples of their application are provided for selected situations. Because some indoor air quality investigations may require application of complex, multistage, survey sampling procedures and because this standard is a guide rather than a practice, the references in Appendix X1 are recommended for guidance on appropriate probability sampling methods, rather than including expositions of such methods in this guide.

1.4 This standard does not address non-probability sampling approaches. Non-probability sampling approaches may be needed, such as worst-case sampling, range finding sampling, and screening sampling as inputs to help guide and inform probability sampling methods.

1.5 *Units*—The values stated in SI units are to be regarded as standard. No other units of measurement are included in this standard.

1.6 *This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

D1356 Terminology Relating to Sampling and Analysis of Atmospheres

3. Terminology

3.1 *Definitions*—For definitions of terms used in this guide, refer to Terminology D1356.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *census, n*—survey of all elements of the target population.

3.2.2 *cluster sample, n*—a sample in which the sampling frame is partitioned into disjoint subsets called clusters and a sample of the clusters is selected.

3.2.2.1 *Discussion*—Data may be collected for all units in each sample cluster or, when a multistage sample is being selected, the units within the sampled clusters may be further subsampled.

3.2.3 *compositing samples, v*—physically combining the material collected in two or more environmental samples.

3.2.4 *expected value, n*—the average value of a sample statistic over all possible samples that could be selected using a specified sample selection procedure.

3.2.5 *multistage sample, n*—a sample selected in stages such that larger units are selected at the first stage, and smaller units are selected at each subsequent stage from within the units selected at the previous stage of sampling.

3.2.5.1 *Discussion*—For assessing the indoor air quality in a population of office buildings, individual buildings might be

[1] This guide is under the jurisdiction of ASTM Committee D22 on Air Quality and is the direct responsibility of Subcommittee D22.05 on Indoor Air.

Current edition approved Sept. 1, 2023. Published September 2023. Originally approved in 1995. Last previous edition approved in 2017 as D5791 – 95 (2017). DOI: 10.1520/D5791-23.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

selected at the first stage of sampling, floors selected within sample buildings at the second stage, and monitoring locations (for example, rooms or grid points) selected on sampled floors at the third stage.

3.2.6 *population parameter, n*—a characteristic based on or calculated from all units in the target population.

3.2.6.1 *Discussion*—The purpose of selecting a sample is usually to estimate population parameters. Population parameters cannot actually be calculated unless data are available for all units in the population.

3.2.7 *probability sample, n*—a sample for which every unit on the sampling frame has a known, positive probability of being selected into the sample.

3.2.7.1 *Discussion*—The terms *probability sampling* and *random sampling* are sometimes used interchangeably.

3.2.8 *sampling frame, n*—a list from which a sample is selected.

3.2.8.1 *Discussion*—An ideal sampling frame contains each member of the target population exactly once and contains no units that are not members of the target population. In practice, the sampling frame may miss some members of the target population (for example, new employees in a building) and include some individuals who are not members of the target population (for example, individuals who no longer work in the building). However, no member of the population should be listed more than once on the sampling frame.

3.2.9 *simple random sample, n*—a sample of *n* elements selected from the sampling frame in such a way that all possible samples of *n* elements have the same chance of being selected.

3.2.10 *statistic, n*—a sample-based estimate of a population parameter.

3.2.11 *stratified sample, n*—a sample in which the sampling frame is partitioned into disjoint subsets called strata, and sample units are selected independently from each stratum, possibly at different sampling rates.

3.2.12 *systematic sample, n*—a sample selected by choosing one of the first *k* elements on the sampling frame at random and then including every *k*th element thereafter.

3.2.13 *target population, n*—the set of units or elements (for example, people or locations in space and time) about which a sample is designed to provide inferences.

3.2.13.1 *Discussion*—The target population is sometimes referred to as the population or universe of interest.

3.2.14 *unbiased estimator, n*—a statistic whose expected value is equal to the population parameter that it is intended to estimate.

4. Summary of Guide

4.1 When the objectives of an investigation of indoor air quality include extending inferences from a sample of units to the larger population from which those units were selected, probability sampling methods must be used to select the sample units to be observed and measured. Examples include:

4.1.1 Estimating the distributions of health and comfort symptoms experienced by the employees in a particular building during a specific week.

4.1.2 Estimating the distribution of hourly average concentrations of specific substances in the breathing zone air in a particular building during the working hours of a specific week.

4.1.3 Estimating the relationship between measures of environmental conditions in a building and the health or comfort symptoms experienced by the occupants.

4.1.4 Thus, the study objectives are always a key consideration for determining if probability sampling methods are necessary. Potential objectives for indoor air studies that would require probability sampling methods are discussed explicitly in Section 6.

4.2 Guidance is provided regarding the appropriate probability sampling methods to address these and other goals that require extending inferences from a sample to a specific population. Those sampling methods require construction of a sampling frame from which population elements can be selected. Examples include:

4.2.1 A list of all offices or work stations in a building,

4.2.2 A grid of potential monitoring locations that effectively covers the entire population of interest, and

4.2.3 A list of all persons who work in a specific building.

4.3 Since environmental concentrations usually vary continuously in time, spatial frame units like those listed in 4.2 often must be crossed with temporal units, such as seasons, weeks, days, or hours, to form sampling frame units (for example, building-seasons, office-weeks, or person-days). Specific issues that must be considered when constructing these types of sampling frames are discussed in Section 7.

4.4 In addition to constructing sampling frames, a randomization procedure is necessary so that units can be selected from the frame with known probabilities. Some basic considerations for and methods of selecting probability samples for studies of indoor air quality are presented in Section 8.

4.5 Finally, Section 9 discusses considerations for statistical analysis and reporting that are peculiar to data collected using probability sampling designs. Special statistical analysis methods are necessary when the sampling design includes stratification, clustering, multistage sampling, or unequal probabilities of selection.

5. Significance and Use

5.1 Studies of indoor air problems are often iterative in nature. A thorough engineering evaluation of a building **(1-4)**[3] is sometimes sufficient to identify likely causes of indoor air problems. When these investigations and subsequent remedial measures are not sufficient to solve a problem, more intensive investigations may be necessary.

5.2 This guide provides the basis for determining when probability sampling methods are needed to achieve statistically defensible inferences regarding the goals of a study of

[3] The boldface numbers in parentheses refer to the list of references at the end of this standard.