Designation: ~~D6044 – 96 (Reapproved 2015)~~ D6044 – 21

Standard Guide for Representative Sampling for Management of Waste and Contaminated Media[1]

This standard is issued under the fixed designation D6044; 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 the definition of representativeness in environmental sampling, identifies sources that can affect representativeness (especially bias), and describes the attributes that a representative sample or a representative set of samples should possess. For convenience, the term “representative sample” is used in this guide to denote both a representative sample and a representative set of samples, unless otherwise qualified in the text.

1.2 This guide outlines a process by which a representative sample may be obtained from a population. The purpose of the representative sample is to provide information about a statistical parameter(s) (such as mean) of the population regarding some characteristic(s) (such as concentration) of its constituent(s) (such as lead). This process includes the following stages: (*1*) minimization of sampling bias and optimization of precision while taking the physical samples, (*2*) minimization of measurement bias and optimization of precision when analyzing the physical samples to obtain data, and (*3*) minimization of statistical bias when making ~~inference~~inferences from the sample data to the population. While both bias and precision are covered in this guide, major emphasis is given to bias reduction.

1.3 This guide describes the attributes of a representative sample and presents a general methodology for obtaining representative samples. It does not, however, provide specific or comprehensive sampling procedures. It is the user’s responsibility to ensure that proper and adequate procedures are used.

1.4 The assessment of the representativeness of a sample is not covered in this guide since it is not possible to ever know the true value of the population.

1.5 Since the purpose of each sampling event is unique, this guide does not attempt to give a ~~step by step~~ step-by-step account of how to develop a sampling design that results in the collection of representative samples.

1.6 Appendix X1 contains two case ~~studies,~~studies which discuss the factors for obtaining representative samples.

1.7 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate ~~safety~~ safety, health, and ~~health~~environmental practices and determine the applicability of regulatory limitations prior to use.*

1.8 *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.*

[1] This guide is under the jurisdiction of ASTM Committee D34 on Waste Management and is the direct responsibility of Subcommittee D34.01.01 on Planning for Sampling.

Current edition approved ~~Sept. 1, 2015~~Oct. 1, 2021. Published ~~September 2015~~October 2021. Originally approved in 1996. Last previous edition approved in ~~2009~~2015 as D6044 – 96 ~~(2009).~~(2015). DOI: ~~10.1520/D6044-96R15.~~10.1520/D6044-21.

2. Referenced Documents

2.1 *ASTM Standards:*[2]

D3370 Practices for Sampling Water from Flowing Process Streams
D4448 Guide for Sampling Ground-Water Monitoring Wells
D4547 Guide for Sampling Waste and Soils for Volatile Organic Compounds
~~D4700 Guide for Soil Sampling from the Vadose Zone~~
D4823 Guide for Core Sampling Submerged, Unconsolidated Sediments
D5088 Practice for Decontamination of Field Equipment Used at Waste Sites
D5681 Terminology for Waste and Waste Management
D5792 Practice for Generation of Environmental Data Related to Waste Management Activities: Development of Data Quality Objectives
D5956 Guide for Sampling Strategies for Heterogeneous Wastes
D6051 Guide for Composite Sampling and Field Subsampling for Environmental Waste Management Activities
D6169 Guide for Selection of Soil and Rock Sampling Devices Used With Drill Rigs for Environmental Investigations
D6286 Guide for Selection of Drilling and Direct Push Methods for Geotechnical and Environmental Subsurface Site Characterization
D6634 Guide for Selection of Purging and Sampling Devices for Groundwater Monitoring Wells
D6771 Practice for Low-Flow Purging and Sampling for Wells and Devices Used for Ground-Water Quality Investigations
D7929 Guide for Selection of Passive Techniques for Sampling Groundwater Monitoring Wells

3. Terminology

~~3.1 *analytical unit, n*—the actual amount of the sample material analyzed in the laboratory.~~

~~3.2 *bias, n*—a systematic positive or negative deviation of the sample or estimated value from the true population value.~~

~~3.2.1 *Discussion*—This guide discusses three sources of bias—sampling bias, measurement bias, and statistical bias.~~

~~There is a sampling bias when the value inherent in the physical samples is systematically different from what is inherent in the population.~~

~~There is a measurement bias when the measurement process produces a sample value systematically different from that inherent in the sample itself, although the physical sample is itself unbiased. Measurement bias can also include any systematic difference between the original sample and the sample analyzed, when the analyzed sample may have been altered due to improper procedures such as improper sample preservation or preparation, or both.~~

~~There is a statistical bias when, in the absence of sampling bias and measurement bias, the statistical procedure produces a biased estimate of the population value.~~

~~Sampling bias is considered the most important factor affecting inference from the samples to the population.~~

~~3.3 *biased sampling, n*—the taking of a sample(s) with prior knowledge that the sampling result will be biased relative to the true value of the population.~~

~~3.3.1 *Discussion*—This is the taking of a sample(s) based on available information or knowledge, especially in terms of visible signs or knowledge of contamination. This kind of sampling is used to detect the presence of localized contamination or to identify the source of a contamination. The sampling results are not intended for generalization to the entire population. This is one form of authoritative sampling (see *judgment sampling*.)~~

~~3.4 *characteristic, n*—a property of items in a sample or population that can be measured, counted, or otherwise observed, such as viscosity, flash point, or concentration.~~

~~3.5 *composite sample, n*—a combination of two or more samples.~~

~~3.6 *constituent, n*—an element, component, or ingredient of the population.~~

[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.

3.6.1 *Discussion*—If a population contains several contaminants (such as acetone, lead, and chromium), these contaminants are called the constituents of the population.

3.7 *Data Quality Objectives, DQOs, n*—qualitative and quantitative statements derived from a DQO process describing the decision rules and the uncertainties of the decision(s) within the context of the problem(s) (see Practice D5792).

3.8 *Data Quality Objective Process*—a quality management tool based on the Scientific Method and developed by the U.S. Environmental Protection Agency to facilitate the planning of environmental data collection activities. The DQO process enables planners to focus their planning efforts by specifying the use of data (the decision), the decision criteria (action level), and the decision maker's acceptable decision error rates. The products of the DQO process are the DQOs (see Practice D5792).

3.9 *error, n*—the random or systematic deviation of the observed sample value from its true value (see *bias* and *sampling error*).

3.10 *heterogeneity, n*—the condition or degree of the population under which all items of the population are not identical with respect to the characteristic(s) of interest.

3.10.1 *Discussion*—Although the ultimate interest is in the statistical parameter such as the mean concentration of a constituent of the population, heterogeneity relates to the presence of differences in the characteristics (for example, concentration) of the units in the population. It is due to the presence of fundamental heterogeneity (or fundamental error)[3] in the population that sampling variance arises. Degree of sampling variance defines the degree of precision in estimating the population parameter using the sample data. The smaller the sampling variance is, the more precise the estimate is. See also *sampling error*.

3.11 *homogeneity, n*—the condition of the population under which all items of the population are identical with respect to the characteristic(s) of interest.

3.12 *judgment sampling, n*—taking of a sample(s) based on judgment that it will more or less represent the average condition of the population.

3.12.1 *Discussion*—The sampling location(s) is selected because it is judged to be representative of the average condition of the population. It can be effective when the population is relatively homogeneous or when the professional judgment is good. It may or may not introduce bias. It is a useful sampling approach when precision is not a concern. This is one form of authoritative sampling (see *biased sampling*.)

3.13 *population, n*—the totality of items or units of materials under consideration.

3.14 *representative sample, n*—a sample collected in such a manner that it reflects one or more characteristics of interest (as defined by the project objectives) of a population from which it is collected.

3.14.1 *Discussion*—A representative sample can be a single sample, a collection of samples, or one or more composite samples. A single sample can be representative only when the population is highly homogeneous.

3.15 *representative sampling, n*—the process of obtaining a representative sample or a representative set of samples.

3.16 *representative set of samples, n*—a set of samples that collectively reflect one or more characteristics of interest of a population from which they were collected. See *representative sample*.

3.17 *sample, n*—a portion of material that is taken for testing or for record purposes.

3.17.1 *Discussion*—Sample is a term with numerous meanings. The scientist collecting physical samples (for example, from a landfill, drum, or monitoring well) or analyzing samples considers a sample to be that unit of the population that was collected and placed in a container. A statistician considers a sample to be a subset of the population, and this subset may consist of one or more physical samples. To minimize confusion, the term *sample*, as used in this guide, is a reference to either a physical sample held in a sample container, or that portion of the population that is subjected to in situ measurements, or a set of physical samples. See *representative sample*.

~~3.17.1.1 The term *sample size* also means different things to the scientist and the statistician. To avoid confusion, terms such as sample mass/sample volume and number of samples are used instead of sample size.~~

3.1 ~~*sampling error*—~~*Definitions—*~~the systematic and random deviations of the sample value from that of the population.~~For definitions of terms used in this standard, refer to Terminology D5681~~The systematic error is the~~ .~~sampling bias. The random error is the sampling variance.~~

~~3.18.1 *Discussion*—Before the physical samples are taken, potential sampling variance comes from the inherent population heterogeneity (sometimes called the "fundamental error," see *heterogeneity*). In the physical sampling stage, additional contributors to sampling variance include random errors in collecting the samples. After the samples are collected, another contributor is the random error in the measurement process. In each of these stages, systematic errors can occur as well, but they are the sources of bias, not sampling variance.~~

~~3.18.1.1 Sampling variance is often used to refer to the total variance from the various sources.~~

~~3.19 *stratum, n*—a subgroup of the population separated in space or time, or both, from the remainder of the population, being internally similar with respect to a target characteristic of interest, and different from adjacent strata of the population.~~

~~3.19.1 *Discussion*—A landfill may display spatially separated strata, such as old cells containing different wastes than new cells. A waste pipe may discharge into temporally separated strata of different constituents or concentrations, or both, if night-shift production varies from the day shift. In this guide, strata refer mostly to the stratification in the concentrations of the same constituent(s).~~

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *analytical unit, n*—the actual amount of the sample material analyzed in the laboratory.

3.2.2 *bias, n*—a systematic positive or negative deviation of the sample or estimated value from the true population value.

3.2.2.1 *Discussion—*

This guide discusses three sources of bias: sampling bias, measurement bias, and statistical bias.

3.2.2.2 *Discussion—*

There is a sampling bias when the value inherent in the physical samples is systematically different from what is inherent in the population.

3.2.2.3 *Discussion—*

There is a measurement bias when the measurement process produces a sample value systematically different from that inherent in the sample itself, although the physical sample is itself unbiased. Measurement bias can also include any systematic difference between the original sample and the sample analyzed, when the analyzed sample may have been altered due to improper procedures such as improper sample preservation or preparation, or both.

3.2.2.4 *Discussion—*

There is a statistical bias when, in the absence of sampling bias and measurement bias, the statistical procedure produces a biased estimate of the population value.

3.2.2.5 *Discussion—*

Sampling bias is considered the most important factor affecting inference from the samples to the population.

3.2.3 *biased sampling, n*—the taking of a sample(s) with prior knowledge that the sampling result will be biased relative to the true value of the population.

3.2.3.1 *Discussion—*

This is the taking of a sample(s) based on available information or knowledge, especially in terms of visible signs or knowledge of contamination. This kind of sampling is used to detect the presence of localized contamination or to identify the source of a contamination. The sampling results are not intended for generalization to the entire population. This is one form of authoritative sampling (see *judgment sampling*).

3.2.4 *composite sample, n*—a combination of two or more samples.

3.2.5 *constituent, n*—an element, component, or ingredient of the population.

3.2.5.1 *Discussion—*
If a population contains several contaminants (such as acetone, lead, and chromium), these contaminants are called the constituents of the population.

3.2.6 *error, n—*the random or systematic deviation of the observed sample value from its true value (see *bias* and *sampling error*).

3.2.7 *homogeneity, n—*the condition of the population under which all items of the population are identical with respect to the characteristic(s) of interest.

3.2.8 *judgment sampling, n—*taking of a sample(s) based on judgment that it will more or less represent the average condition of the population.

3.2.8.1 *Discussion—*
The sampling location(s) is selected because it is judged to be representative of the average condition of the population. It can be effective when the population is relatively homogeneous or when the professional judgment is good. It may or may not introduce bias. It is a useful sampling approach when precision is not a concern. This is one form of authoritative sampling (see *biased sampling*).

3.2.9 *representative sample, n—*a sample collected in such a manner that it reflects one or more characteristics of interest (as defined by the project objectives) of a population from which it is collected.

3.2.9.1 *Discussion—*
A representative sample can be a single sample, a collection of samples, or one or more composite samples. A single sample can be representative only when the population is highly homogeneous.

3.2.10 *sample, n—*one or more items or portions collected from a lot or population.

3.2.10.1 *Discussion—*
Sample is a term with numerous meanings. The scientist collecting physical samples (for example, from a landfill, drum, or monitoring well) or analyzing samples considers a sample to be that unit of the population that was collected and placed in a container. A statistician considers a sample to be a subset of the population, and this subset may consist of one or more physical samples. To minimize confusion, the term *sample*, as used in this guide, is a reference to either a physical sample held in a sample container, or that portion of the population that is subjected to in-situ measurements, or a set of physical samples. See *representative sample*.

3.2.10.2 *Discussion—*
The term *sample size* also means different things to the scientist and the statistician. To avoid confusion, terms such as sample mass/sample volume and number of samples are used instead of sample size.

3.2.11 *sampling error, n—*the systematic and random deviations of the sample value from that of the population. The systematic error is the *sampling bias*. The random error is the *sampling variance*.

3.2.11.1 *Discussion—*
Before the physical samples are taken, potential sampling variance comes from the inherent population heterogeneity (sometimes called the "fundamental error"; see *heterogeneity*). In the physical sampling stage, additional contributors to sampling variance include random errors in collecting the samples. After the samples are collected, another contributor is the random error in the measurement process. In each of these stages, systematic errors can occur as well, but they are the sources of bias, not sampling variance.

3.2.11.2 *Discussion—*
Sampling variance is often used to refer to the total variance from the various sources.

3.2.12 *stratum, n—*a subgroup of the population separated in space or time, or both, from the remainder of the population, being internally similar with respect to a target characteristic of interest, and different from adjacent strata of the population.

3.2.12.1 *Discussion—*
A landfill may display spatially separated strata, such as old cells containing different wastes than new cells. A waste pipe may discharge into temporally separated strata of different constituents or concentrations, or both, if night-shift production varies from the day shift. In this guide, strata refer mostly to the stratification in the concentrations of the same constituent(s).

~~3.20 *subsample, n—*a portion of the original sample that is taken for testing or for record purposes.~~

4. Significance and Use

~~4.1 Representative samples are defined in the context of the study objectives.~~

4.1 This guide defines the meaning of a representative sample, as well as the attributes the sample(s) needs to have in order to provide a valid inference from the sample data to the population.

4.2 This guide also provides a process to identify the sources of error (both systematic and random) so that an effort can be made to control or minimize these errors. These sources include sampling error, measurement error, and statistical bias.

4.3 When the objective is limited to the taking of a representative (physical) sample or a representative set of (physical) samples, only potential sampling errors need to be considered. When the objective is to make an inference from the sample data to the population, additional measurement error and statistical bias need to be considered.

4.4 This guide does not apply to the cases where the taking of a nonrepresentative sample(s) is prescribed by the study objective. In that case, sampling approaches such as judgment sampling or biased sampling can be taken. These approaches are not within the scope of this guide.

4.5 Following this guide does not guarantee that representative samples will be obtained. But failure to follow this guide will likely result in obtaining sample data that are either biased or imprecise, or both. Following this guide should increase the level of confidence in making the inference from the sample data to the population.

4.6 This guide can be used in conjunction with the DQO process (see Practice D5792).

4.7 This guide is intended for those who manage, design, and implement sampling and analytical plans for waste management and contaminated media.

5. Representative Samples

5.1 Samples are taken to ~~infer~~ make inferences about some statistical parameter(s) of the population regarding some characteristic(s) of its constituent(s) of interest. This is discussed in the following sections.

5.2 *Samples*—When a representative sample consists of a single physical sample, it is a sample that by itself reflects the characteristics of interest of the population. On the other hand, when a representative sample consists of a set of physical samples, the samples collectively reflect some characteristics of the population, though the samples individually may not be representative. In most cases, more than one physical sample is necessary to characterize the population, because the population in environmental sampling is usually heterogeneous.

5.3 *Constituents and Characteristics*—A population can possess many constituents, each with many characteristics. Usually it is only a subset of these constituents and characteristics that are of interest in the context of the stated problem. Therefore, samples need to be representative of the population only in terms of these constituent(s) and characteristic(s) of interest. A sampling plan needs to be designed accordingly.

5.4 *Parameters*—Similarly, samples need to be representative of the population only in the parameter(s) of interest. If the interest is only in estimating a parameter such as the population mean, then composite samples, when taken correctly, will not be biased and therefore constitute a representative sample (regarding bias) for that parameter. On the other hand, if the interest happens to be the estimation of the population variance (of individual sampling units), another parameter, then the variance of the composite samples is a biased estimate of the population variance and therefore is not representative. (It is to be noted that composite samples are often used to increase the precision in estimating the population mean and not to estimate the population variance of individual sampling units.)

5.5 *Population*—Since the samples are intended to be representative of a population, a population must be well defined, especially in its spatial or temporal boundaries, or both, according to the study objective.

5.6 *Representativeness*—The word "reflects" in this guide is used to mean a certain degree of low bias and high precision when comparing the sample value(s) to the population value(s). This is a broad definition of sample representativeness used in this guide. A narrower definition of representativeness is often used to mean simply the absence of bias.

5.6.1 *Bias*—Bias is sometimes mistakenly taken to be "a difference between the observed value of a physical sample and the true population value." The correct definition of bias is "a *systematic* (or consistent) difference between an observed (sample) value and the true population value." The word "systematic" here implies "on the average" over a set of physical samples, and not a single physical sample. Recall that sampling error consists of the random and systematic deviations of a sample (or estimated) value from that of the population. Although random deviations may occur on occasions due to imprecision in the sampling or measurement processes, or both, they balance out on the average and lead to no systematic difference between the sample (or estimated) value and the population value. The random deviation corresponds to the observation of "a random difference between a single physical sample value and the true population value," which can be randomly positive or negative, and is not a bias. On the other hand, a persistent positive or negative difference is a systematic error and is a bias.

5.6.1.1 In order to assess bias, the true population value must be known. Since the true population value is rarely known, bias cannot be quantitatively assessed. However, this guide provides an approach to identifying the potential sources of bias and general considerations for controlling or minimizing these potential biases.

5.6.2 *Precision*—Precision has to do with the level of confidence in estimating the population value using the sample data. If the population is totally homogeneous and the measurement process is flawless, a single sample will provide a completely precise estimate of the population value. When the population is heterogeneous or the measurement process is not totally precise, or both, a larger number of samples will provide a more precise estimate than a smaller number of samples.

5.6.2.1 In the case of bias, the goal in environmental sampling is its absence. In the case of precision, the goal in sampling will depend on factors such as:

(1) The precision level needed to achieve the desired levels of decision errors, both false positive and false negative errors,

(2) If the true value is known or suspected to be well below the regulatory limit, high precision in the samples may not be needed, and

(3) The study budget.

5.6.2.2 Note that the second item applies similarly to bias as well.

5.6.2.3 Since bias, especially during sampling, can be very large when proper procedures are not followed, it is considered to be the first necessary condition for sample representativeness. On the other hand, precision can be more or less controlled, for example, by increasing the number of samples taken or by decreasing the sampling or measurement variabilities, or both.

5.6.2.4 The optimal number of samples to take to achieve a desired level of precision is typically an issue in optimization of a sampling plan. Therefore, the precision issue will be covered only briefly in this guide.

6. A Systematic Approach to Representative Sampling

6.1 A systematic approach is one that first defines the desired end result and then designs a process by which such a result can be obtained. In representative sampling, the desired end result is a sample or a set of samples that achieves desired levels of low bias and high precision.

6.2 A representative sampling process is described in Fig. 1. The key components in the process are described in this section.

6.3 *Study Objective*—A sampling plan is designed according to a defined problem or a stated study objective. The samples are then collected according to the sampling plan. Generally, the study objective dictates that representative samples be taken for the purpose of inference about the population. In that case, these samples will need to be collected according to this guide in order for the inference to be valid. Occasionally, the objective is merely to detect the presence of a contaminant or to obtain a "worst case" sample. In that case, an authoritative sampling approach (biased sampling or judgment sampling) may be taken and this guide does not apply.

6.4 *Population*—A population consists of the totality of items or units of materials under consideration (Compilation of ASTM