A Methodology for Assessing Sample Representativeness

Charles A. Ramsey,¹ and Alan D. Hewitt²

¹EnviroStat, Inc., Fort Collins, CO, USA
²U.S. Army Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory, Hanover, NH, USA

Assessing sample representativeness is a critical component of any environmental investigation and should be performed before any conclusions are reached. If the samples are not representative, any conclusions or decisions will be incorrect. A complete understanding of the data quality objective process, sample plan design, sample plan implementation, and quality control is required to assess sample representativeness. This article presents a methodology for the evaluation of sample representativeness.

Keywords: representativeness, quality control, data quality objectives, sample plan design

Introduction

The term “representative” has many definitions. A variety of definitions can be found elsewhere (Warren, 2004). However, a wide variety of definitions is not helpful since representative is an operational definition. If a definition cannot be agreed on, it is impossible to evaluate representativeness in a meaningful way. If evidence for representativeness is not presented, the data cannot be characterized as effective for project decision-making (Crumbling, 2001). Perhaps Deming (1986) is the most enlightened on this topic when he addresses the concept of representativeness in his book Out of the Crisis:

What is the meaning of the adjective representative?

Answer: the word has no meaning.

The word representative is an adjective and thus it must be asked “representative of what?” Assessing representativeness can only be accomplished in the context of the question the data are supposed to address. In the simplest terms, if the data can answer the question, it is representative.

There are many aspects of representative data:

- Are the field samples representative of the field (target) population?
- Is the subsampled material in the laboratory representative of the field sample?
- Were all the sample preparation steps performed in a way not to destroy the integrity (representativeness) of the sample?
- Was the chemical analysis performed in such a way to yield representative analytical results of the subsampled material?

This article will discuss assessing the first two aspects of data representativeness—field sampling and laboratory subsampling—since these are the major sources of error in environmental investigations (Gy, 1998; Jenkins et al., 1997; Pitard, 2000; USEPA, 2003; Wait, 2002; Walsh et al., 2002). This article will not address aspects of sampling beyond the sample selection process. Any activities beyond actual sample selection (e.g., container, type, preservation) would be considered as part of sample preparation.

Representativeness of samples (field and laboratory) is a function of sampling error. If no sampling error exists, the samples would by definition be representative. There is, however, always some degree of sampling error unless the entire population is analyzed. Analysis of the entire population of interest is typically cost prohibitive or impractical. Thus, determination of sampling error is a critical component of assessing sample representativeness.

Assessment of Field Sample Representativeness

Components of Assessing Sample Representativeness

The three major components of assessing sample representativeness are:

1. Determining the data quality objectives
2. Evaluating sample plan design
3. Evaluating specific quality control data

Determining the Data Quality Objectives

The Environmental Protection Agency (EPA) Data Quality Objective (DQO) process is a seven-step process that is employed to develop a scientific plan for data collection (USEPA, 2000). The DQO process or other type of systematic planning must be used if an environmental investigation is to have scientific merit. For the purposes of this article, only three steps of the EPA DQO...
process are addressed, as these steps are critical to designing a sampling plan and assessing sample representativeness. The three steps are:

1. What is the question?
2. What population is the question referring to?
3. What is the desired confidence in the answer to the question?

What Is the Question?
This step asks a question about a population to be sampled. There are many questions that could be asked about a population. Some common examples are:

1. What is the average concentration?
2. Are there trends with time or space?
3. Is the concentration below a limit?
4. Were the treatment standards met?

What Is the Population?
This step defines the population that is to be sampled. There are many possible populations in an environmental investigation. Careful selection of the population is critical to the success of an environmental investigation (Gilbert, 1987). In many cases, this selection is not clearly defined.

The population is the material that the question refers to. The population may be called the “decision unit,” since it is the material involved in the decision. It is also called the target population, management unit, remedial unit, exposure unit, and so forth. Definition of the population is typically the most difficult and critical part of the DQO process.

The selection of the population is critical to determine where the samples will be collected from and where the inferences will be made to. The choice of the population also determines what sampling tools should be used. If not all the population is available (accessible for sampling), it will limit the inferences that can be made.

The population determination must be very precise. An example of a common population is soil in a yard. But this is not precise enough. How is soil defined? Different professions define it differently. A geologist and an agronomist have differing definitions. Soil may be defined as “passing a 2-mm sieve.” This is a workable definition. How is the yard defined? Is it the legal description, the fenced area, the area generally used (not the soil under the stairs or the rose bushes)? How deep does soil go? The top inch, two inches, or six inches? Without an exact definition it is impossible to know where to collect the samples from, where the inferences will be made to, and what type of tools should be used.

Another example is a collection of drums. Is each single drum a population or are all the drums a single population? The sample plan design and conclusions are different for each situation.

What Is the Desired Confidence?
This step defines the confidence with which sample(s) (or, in a broader sense, the data) can answer the question about the population. There are two parts to defining the confidence. One is what type of confidence is required and the other is the level of confidence required. There are two possible types of confidence, statistical and professional. Statistical confidence is used when a confidence or probability statement (e.g., 90% confidence, the mean is below the limit) is desired. Professional confidence is employed when a conclusion is based on professional judgment (e.g., in my professional opinion the mean is below the limit). Either type of confidence is acceptable depending on the DQOs of the study.

If statistical confidence is required, some type of random selection of samples from the population is required. When professional confidence is utilized, the samples can be selected randomly or based on judgment.

The objective of statistical inference is to make claims about the entire population based on the analysis of a few grams of material from the population. In other words, a statement may be made regarding literally tons of material on the basis of the results from a few grams of material. In order to make such an inference there are several requirements:

- First of all, the entire population must be available for sample collection. However, in some cases the entire population is not available. If so, statistical inferences can only be made to the available population. For example, only the top portion (12 inches) of a drum (containing 36 inches of material) may be available because the bottom (24 inches) of material cannot be reached with the sampling technique employed. Even if the sampling and analysis is perfect (no errors), any statistical claim can only be applied to the top (available) portion of the drum (target population). If this claim is not adequate to meet the DQOs, the sampling strategy will have to be changed to allow access to the entire target population.
- Second, the material must be collected in some type of random fashion.
- Third, the sampling must be performed to adequately represent the material (see Sample Plan Design).

The desired level of confidence drives the amount of effort and quality control required by the sampling plan design. Typically, a higher level of effort is required for a higher level of statistical confidence. For professional confidence, the person making the claim determines the adequacy of the sampling, quality control, etc. In other words, the person making the claim evaluates whether data are sufficient or representative to help support their claim.

Sample Plan Design
The DQOs address the question “representative of what.” The sample plan design must be evaluated to determine if the samples collected are representative of and sufficient to answer the question about the population with the desired confidence. It is critical that the design be based on sound scientific principles. Developing a sampling plan on the basis of what has always been done, what you can get away with, and so on, is not use of sound
scientific principles. This will lead to decisions that are neither correct nor defensible.

Sound scientific sampling plan design is based on the compositional and distributional heterogeneity of the material that makes up the population. Heterogeneity is manifested as variability, which can be quite large (Feenstra, 2003). The details of heterogeneity are not addressed in this article but are vital to the collection of a sample that represents a population (Pitard, 2000). In simple terms the sample should be a miniature of the target population. The sample should contain some of each and every type of particle/molecule that makes up the target population in the same proportion that exists in the target population. The target population may have both spatial and temporal boundaries.

**Quality Control**

With any design, including a sampling plan, the design validity must be checked. Quality control is used to check the sample plan design. If a sampling plan is designed to represent a population, quality control can be implemented to test that hypothesis. If the sampling plan is valid, the quality control data should support it. This article does not address all aspects of quality assurance and control, it only addresses quality control that is used to validate a sample plan design.

A scientifically designed sampling plan is based on tolerable error. There are three types of possible error: blunders, systematic errors, and random errors (Maney, 2002). Blunders are not addressed in this article. The measurement of systematic errors in sample selection is very difficult, if not impossible, and can only be controlled by correct selection and use of sampling tools (Pitard, 2000). There is no practical quality control to measure systematic errors in sample selection; therefore, correct selection and use of sampling tools is critical. The magnitude of random errors can be determined from replication (e.g., duplicates) of the sample design. If the sample design produces a representative sample, replication (collection of another sample under the same conditions) would produce another representative sample. If the sample design is to take material from X random locations, replication would be to collect material from X different random locations. Replication is not shifting the sample collection point(s) a few inches and collecting another sample.

Quality control provides information that a design is invalid, but it may not be conclusive to demonstrate that a design is valid. For instance, a sampling plan may fail to detect “hot spots” in a population and thus underestimate the true concentration. Evaluation of the quality control data might indicate the duplicate results are within the tolerable error of the design, but that does not mean that the samples were representative of the population, since both samples may have missed the “hot spot.” Conversely, if the replicate data disagree (error greater than the design criteria) it can be inferred that the samples are not representative of the target population.

The type, quantity, and interpretation of quality control should be dictated by the data quality objective process and not arbitrarily specified. In many cases the same quality control samples and desired control limits are carried forward from work plan to work plan. When this occurs it is evident that the data quality objectives are not being correctly implemented. In most cases the control limits are specified in advance, and if the control limits are met the data is deemed acceptable for decision-making purposes. Following is an example where specifying the quality control in advance of sample collection is a dangerous practice. Random error allowed is specified as 35% in advance of sample collection. If the random error of the samples is less than 35%, the data are deemed acceptable for decision-making purposes. If the random error is greater than 35%, the data fail to meet the specified quality control criteria and are unacceptable for decision-making purposes. However, if the decision limit is 10 mg/L and the population concentration is 12 mg/L, a 35% random error is unacceptable to make a decision. If the population concentration were 2 mg/L, a larger error of 50% would be acceptable. Thus, the specification of the tolerable error prior to a legitimate assessment of allowable error is a dangerous practice.

In summary, quality control is used to measure the random sampling error of the sample plan design. If this error is less than the tolerable error allowed by the DQO process and less than the sampling error the sample plan was designed around, the samples are representative. Systematic errors cannot be measured with any type of quality control, so great care has to be taken with sampling tool selection and use.

**Steps in Assessment of Sample Representativeness**

*Determine the Data Quality Objectives*

Are the question, population, and confidence clearly defined? If not, the evaluation of representativeness cannot continue. If the data were collected without consideration of the data quality objectives, the data are likely worthless for decision-making purposes.

*Evaluate Sampling Plan Design*

Was sound scientific sampling theory used in the design of the sampling plan? Did the sampling plan address both the compositional and distributional heterogeneity of the material? If the plan is not scientifically designed, sample representativeness cannot be evaluated. While it is possible that the samples are representative by luck alone, the results cannot be scientifically defended. For example, a design to estimate the average contaminant concentration of a grid based on the concentration of the contaminant of a discrete core from the center of the grid produces a nonrepresentative sample (Jenkins et al., 1999).

*Evaluate Tool Selection and Use*

Were the correct sampling tools used properly? Did the tools reach all parts of the available population? Did all portions of the population have the same probability of selection? If the answer to any of these questions is no, a bias is introduced that
will render the samples nonrepresentative. For example, a coring device may be used to core the top 12 inches (target population) of a sandy material. But if the lower-most portion of the material falls out of the coring device when it is removed from the target population, a bias results.

**Evaluate Randomness and Accessibility**

If statistical confidence was required, was the sample collection random? If not, no statistical inferences can be made. Was the entire target population available for sample collection? If not, the inference must only include the available population; statistical inference cannot be made to the entire target population. For example, a large pile is the target population. If only the surface of the pile is sampled since that is all the sampler or sample tool can reach, statistical inference can only be made to the surface of the pile. Statistical inference to the entire pile is invalid.

**Evaluate Quality Control Data**

Is there replication to determine errors of precision? Was the entire sampling scheme properly replicated? If not, there is no way to evaluate the random errors associated with the design and sample representativeness cannot be determined. Are the errors within the range of errors expected from the sample plan design? If not, the assumptions that went into the design were not valid and the samples are not representative. For example, if the sample plan design (based on the DQOs) is to yield samples that have a sampling error of less than 20% and the replicate data yields a sampling error of 75%, the design or the assumptions that went into the design are not valid; the sample plan design did not yield representative samples and the data will not satisfy the data quality objectives.

**Final Evaluation of Field Sample Representativeness**

If the plan does not pass any of the above criteria, the samples are not representative. Some of the criteria are pass/fail and others are compared to a standard, but all are critical. If the samples are not representative for any reason, decisions cannot be made and the process must begin again. Statistical calculations cannot be used to evaluate sample representativeness since the samples are not representative. This is why correct implementation of the data quality objective process, scientifically designed sampling plans, and proper quality control are critical to project success.

**Assessment of Laboratory Subsample Representativeness**

Sampling in the laboratory, commonly called laboratory subsampling, has the same issues as field sampling (Ramsey and Suggs, 2001). If the subsampling is not performed in a representative manner, a representative field sample will be rendered nonrepresentative in the laboratory and yield nonrepresentative data resulting in incorrect decisions (Walsh et al., 2002).

Laboratory subsample representativeness is evaluated in the same manner as the field samples. The only difference between the field and the laboratory is the target population, which is now the field sample. Therefore, the same sampling errors, sample design, and quality control issues exist, but on a different population. In the laboratory there is no excuse for not having the entire population available; yet that is a major source of nonrepresentativeness in the laboratory. The common laboratory practices of scooping a few grams of material for extraction or stirring the top portion of the field sample prior to scooping a few grams of material for extraction will not yield a representative subsample (USEPA, 2003). Subsampling must have a scientifically based sampling procedure based on DQOs with proper quality control as used in the field.

**Conclusions**

A representative sample is one that answers a question about a population with a certain confidence. Only when the items addressed in this article are clearly and precisely defined can the representativeness of a sample be determined. A sample that is representative for a specific question is most likely not representative for a different question.

In order to collect representative samples, the data quality objectives must be stated very precisely. Failure of many environmental investigations can be traced to lack of specific objectives. Without specific numeric objectives, sampling plans to collect representative samples cannot be designed and evaluation of representativeness is impossible. If data quality objectives are correctly implemented, the assessment of representativeness can be achieved.

Determination of the representativeness of samples is not a matter of statistical analysis of the data after the fact. It is a result of careful planning and proper design. Valid statistical analysis can only be performed once the representativeness of the data is demonstrated—and not before!

**Acknowledgement**

Support was received from U.S. Army Corps of Engineers program “Characterization, Evaluation, and Remediation of Distributed Source Compounds (UXO-C) on Army Ranges.” The Technical Director and Program Manager for this work is Dr. John M. Cullinane, U.S. Army Engineer Research and Development Center, Environmental Laboratory, Vicksburg, MS, USA.

**References**


Implications for source zone mass estimates. Environmental & Engineer-
ing Geoscience v. IX (1) 19–24.
Gilbert, R. O. 1987. Statistical Methods for Environmental Pollution Mon-
Jenkins, T. F., Grant, C. L., Brar, G. S., Thorne, P. G., Schumacher, P.
and analysis of soil samples at TNT-contaminated sites. Field Analytical
Chemistry and Technology 1(3):151–163.
Jenkins, T. F., Grant, C. L., Walsh, M. E., Thorne, P. G., Thiboutot, S.,
effects on sampling and analysis at an HMX-contaminated antitank firing
Science and Technology 31(19):383A–389A.
through scientific subsampling techniques. Environmental Testing &
Analysis 10(2):12–16.
USEPA (U. S. Environmental Protection Agency). 2000. Guidance for the
data quality objectives process QA/G-4. United States Environmental Pro-
USEPA (U. S. Environmental Protection Agency). 2003. Guidance for ob-
taining representative laboratory analytical subsamples from particulate
laboratory samples. United States Environmental Protection Agency,
Washington, DC. EPA/600/R-03/027.
Wait, A. D. 2002. Challenges in producing defensible environmental chem-
415–454.
particle size reduction by grinding on subsampling variance for explosives
Warren, J. 2004. Representativeness at the Environmental Protection