Designation: D6842 – 02 (Reapproved 2007)

StandardGuide for Designing Cost-Effective Sampling and Measurement Plans by Use of Estimated Uncertainty and Its Components in Waste Management Decision-Making[1]

This standard is issued under the fixed designation D6842; 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 Waste management decisions generally involve uncertainty because of the fact that decisions are based on the use of sample data. When uncertainty can be reduced or controlled, a better decision can be achieved. One way to reduce or control uncertainty is through the estimation and control of the components contributing to the overall uncertainty (or variance). Control of the sizes of these variance components is an optimization process. The optimizations results can be used to either improve an existing sampling and analysis plan (if it should be found to be inadequate for decision-making purposes) or to optimize a new plan by directing resources to where the overall variance can be reduced the most.

1.2 Estimation of the variance components from the total variance starts with the sampling and measurement process. The process involves two different kinds of uncertainties: random and systematic. The former is associated with imprecision of the data, while the latter is associated with bias of the data. This guide will discuss only sources of uncertainty of a random nature.

1.3 There may be many sources of uncertainty in waste management decisions. However, this guide does not intend to address the issue of how these sources are identified. It is the responsibility of the stakeholders and their technical staff to analyze the sampling and measurement processes in order to identify the potentially significant sources of uncertainty. After identifying these sources, this guide will provide guidance on how to collect and analyze data to obtain an estimate of the total uncertainty and its components.

2. Terminology

2.1 *analysis of variance (ANOVA), n*—a statistical method of decomposing (or breaking down) the total variance and estimating or testing its contributing component variances for statistical significance.

2.2 *balanced design, n*—a statistical study where replication in each of the levels of ANOVA is identical.

2.3 *measurement process, n*—the method and procedure of obtaining and measuring samples or their subsamples to produce sample data.

2.4 *sampling process, n*—the method and procedure of collecting physical samples from a defined population.

2.5 *unbalanced design, n*—a statistical study where replication in some or all of the levels of ANOVA is not identical.

3. Significance and Use

3.1 This guide will evaluate sample data that contain a high level of uncertainty for decision-making purposes and, where it is feasible, design a statistical study to estimate and reduce the sources of uncertainty. Oftentimes, historical data may be available and adequate for this purpose and no new study is needed.

3.1.1 This approach will help the stakeholders better understand where the greatest sources of uncertainty are in the sampling and analysis process. Resources can be directed to where they can most reduce the overall uncertainty.

3.1.2 Sampling and analysis design under this approach can often be cost-efficient because (*a*) the reduction in uncertainty can be done by statistical means alone and (*b*) the reduction can be translated into a lower number of analyses.

3.2 This guide is limited to the situation where a decision is based on the mean of a population. It will only include discussions of a balanced design for the collection and analysis of sample data in order to estimate the sources of uncertainty. References to unbalanced designs are provided where appropriate.

4. Uncertainty in Decision-Making

4.1 *Decision-Making Based on Data:*

4.1.1 When waste management decision-making is based on data and when the data come from a subset of a population, the data can be used to calculate quantities such as mean, median, or percentage for the purpose of estimating the true value of

[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 Feb. 1, 2007. Published June 2007. Originally approved in 2002. Last previous edition approved in 2002 as D6842 - $02^{\varepsilon1}$. DOI: 10.1520/D6842-02R07.

these quantities in the population. These estimates can be used to make conclusions or decisions about the population on issues such as: (*1*) Is the average concentration of a contaminant at a certain site higher or lower than a regulatory standard? (*2*) Has the cleanup standard been met?

4.1.2 However, these estimates involve uncertainty because of uncertainties in the sampling and measurement processes. The total uncertainty associated with an estimate can be derived from the sample data and it is usually expressed as the variance or standard deviation of the estimate. The estimate and its variance can be used to define the level of confidence in decision-making. For example, they can be used to calculate the upper and lower confidence limits, where the width of the confidence limits is a measure of uncertainty in decision-making.

4.1.3 An example of high data uncertainty and low confidence in decision-making can occur when the sample mean concentration of a site is substantially below a regulatory limit while its upper confidence limit is higher than the regulatory limit. In this case, a reduction in uncertainty will lead to better decision-making. That is, there is a higher probability that the correct decision about the true concentration can be reached and the appropriate action taken.

4.2 *Sampling and Measurement Process:*

4.2.1 When the confidence level is not at the level desired by the decision-makers, the data from the sampling and measurement processes can be analyzed to identify significant sources of contributors to the total variance. This guide will permit project managers to focus on the large sources of uncertainty and allocate resources for their reduction. That, in turn, will improve the sampling and measurement processes and achieve a higher level of confidence in project decisions.

4.2.2 This guide is limited to the situation when a decision needs to be made regarding the mean of a population.

4.2.3 This guide is also limited to the discussions of a balanced design for the collection and analysis of sample data in order to estimate the sources of uncertainty. An example of a balanced design is given in Table 1. In Table 1, the letter "*m*" indicates the number of subsamples taken from a field sample and the letter "*k*" indicates the number of replicate analyses performed on each subsample. Note that there is an equal number of subsamples for each of the field samples and an equal number of replicate analyses for each of the subsamples in Table 1. It is this equality in replication at the subsampling level and at the replicate analysis level that constitutes a balanced design. When there is inequality at any of the levels, it is called an unbalanced design. References to unbalanced designs will be provided where appropriate.

4.2.4 A typical sampling and measurement process goes through three stages:

4.2.4.1 The collection of field samples,

4.2.4.2 Taking of subsamples from the field samples in the laboratory, and

4.2.4.3 Duplicate analysis of the subsamples.

4.2.5 The variances associated with each of these stages are known as the sampling variance, subsampling variance, and analytical variance, respectively. The sum of these variances constitutes the total variance in decision-making. The total variance and its contributing components can be estimated from the data when the sampling and measurement process is designed for such purposes. For this guide, the 3-stage sampling and measurement process above will be used as a model for discussion purposes. When other processes are appropriate, consult a statistician.

5. Estimation of Total Variance and Its Components

5.1 *Study Design and Example Data:*

5.1.1 Under any sampling and measurement process, the total variance and its components can be estimated only when the data are collected according to a design. In particular, for the 3-stage process described in 4.2, the variances can be estimated only when there are multiple field samples, where multiple subsamples are taken from each of the field samples and when each of the subsamples is in turn analyzed in multiple replicates (duplicate, triplicate, etc.). The word "multiple" here implies two or more, with two being the minimum requirement. The optimal numbers of field samples, subsamples and replicates will depend on the sizes of their respective variance components and the costs associated with the collection or analysis regarding these components. When the costs are negligible, then they will depend solely on the relative sizes of the variance components alone.

5.1.2 An example of such a study design may appear as noted in Table 1. Example data of TPH concentrations collected from a hypothetical site may appear as shown in Table 2, with the addition of the last 3 columns for the statistical method Analysis of Variance (ANOVA). Note that the data in Table 2 is a balanced design in that the number of subsamples per field sample is equal at 2 and the number of replicate analyses per subsample is equal at 3.

5.1.3 An unbalanced design occurs when the number of subsamples is not equal among the field samples or when the number of replicates is not equal among the subsamples. In this case, the estimation of the variance components becomes more complicated. In this situation, consult a statistician. Some

TABLE 1 Study Design for the Example Sampling and Measurement Process Described in Section 5[A]

Field Sample No.	Subsample No.	Replicate No.	Value
1	1	1	X_{111}
		n	X_{11n}
	m	1	X_{1m1}
		n	X_{1mn}
f	1	1	X_{f11}
		n	X_{f1n}
	m	1	X_{fm1}
		n	X_{fmn}

[A] *f*, *m*, $n \geq 2$ in order to estimate the variance components..

TABLE 2 Example Data of TPH (ppm) for a 3-Stage Sampling and Measurement Process

Field Sample	Sub-sample	TPH in Replicate 1	TPH in Replicate 2	TPH in Replicate 3	Subsample Total	Field Sample Total	Grand Total
1	1	10	11	11	32		
	2	8	7	8	23	55	
2	1	5	6	5	16		
	2	4	4	6	14	30	85