Standard Guide for
Data Assessment for Environmental Waste Management Activities[1]

1. Scope

1.1 This guide covers a practical strategy for examining an environmental project data collection effort and the resulting data to determine if they will support the intended use. It covers the review of project activities to determine conformance with the project plan and impact on data usability. This guide also leads the user through a logical sequence to determine which statistical protocols should be applied to the data.

1.1.1 This guide does not establish criteria for the acceptance or use of data but instructs the assessor/user to use the criteria established by the project team during the planning (data quality objective process), and optimization and implementation (sampling and analysis plan) process.

1.2 The values stated in SI units are to be regarded as standard. No other units of measurement are included in this standard.

1.3 *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 and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

D4687 Guide for General Planning of Waste Sampling

D5088 Practice for Decontamination of Field Equipment Used at Waste Sites

D5283 Practice for Generation of Environmental Data Related to Waste Management Activities: Quality Assurance and Quality Control Planning and Implementation

D5792 Practice for Generation of Environmental Data Related to Waste Management Activities: Development of Data Quality Objectives

3. Terminology

3.1 *Definitions of Terms Specific to This Standard:*

3.1.1 *bias, n*—a systematic error that is consistently negative or consistently positive.

3.1.2 *characteristic, n*—a property of items in a sample or population which can be measured, counted, or otherwise observed.

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

3.1.4 *confidence limit, n*—an upper and/or lower limit(s) within which the true value is likely to be contained with a stated probability or confidence.

3.1.5 *continuous data, n*—data where the values of the individual samples may vary from minus infinity to plus infinity.

3.1.6 *data quality objectives (DQOs), n*—DQOs are qualitative and quantitative statements derived from the DQO process describing the decision rules and the uncertainties of the decision(s) within the context of the problem(s).

3.1.7 *data quality objective process, n*—a quality management tool based on the scientific method and developed to facilitate the planning of environmental data collection activities.

3.1.8 *discrete data, n*—data made up of sample results that are expressed as a simple pass/fail, yes/no, or positive/negative.

3.1.9 *heterogeneity, n*—the condition of the population under which all items of the population are not identical with respect to the parameter of interest.

3.1.10 *homogeneity, n*—the condition of the population under which all items of the population are identical with respect to the parameter of interest.

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

3.1.12 *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.

3.1.13 *sample, n*—a portion of material which is taken from a larger quantity for the purpose of estimating properties or composition of the larger quantity.

3.1.14 *sampling design error, n*—error which results from the unavoidable limitations faced when media with inherently variable qualities are measured and incorrect judgement on the part of the project team.

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

4. Significance and Use

4.1 This guide presents a logical process for determining the usability of environmental data for decision making activities. The process describes a series of steps to determine if the enviromental data were collected as planned by the project team and to determine if the *a priori* expectations/assumptions of the team were met.

4.2 This guide identifies the technical issues pertinent to the integrity of the environmental sample collection and analysis process. It guides the data assessor and data user about the appropriate action to take when data fail to meet acceptable standards of quality and reliability.

4.3 The guide discusses, in practical terms, the proper application of statistical procedures to evaluate the database. It emphasizes the major issues to be considered and provides references to more thorough statistical treatments for those users involved in detailed statistical assessments.

4.4 This guide is intended for those who are responsible for making decisions about environmental waste management projects.

5. General Considerations

5.1 This guide provides general guidance about applying numerical and other techniques to the assessment of data resulting form environmental data collection activities associated with waste management activities.

5.2 The environmental measurement process is a complex process requiring input from a variety of personnel to properly address the numerous issues related to the integrity of the sample collection and measurement process in sufficient detail. Table 1 lists many of the topics that are common to most environmental projects. A well-executed project planning activity (see Guide D4687, Practices D5088, D5283, and D5792) should consider the impact of each of these issues on the reliability of the final project decision. The data assessment process must then evaluate the actual performance in these areas versus that expected by the project planners. Significant deviations from the *a priori* performance level of any one or combination of these issues may impact the reliability of the project decision and necessitate a reconsideration of the decision criteria by the project decision makers.

TABLE 1 Information Needed to Evaluate the Integrity of the Environmental Sample Collection and Analysis Process

General Project Details	
•	Site History
•	Process Description
•	Waste Generation Records
•	Waste Handling/Disposal Practices
•	Sources of Contamination
•	Conceptual Site Model
•	Potential Contaminants of Concern
•	Fate and Transport Mechanisms
•	Exposure Pathways
•	Boundaries of the Study Area
•	Adjacent Properties
Sampling Issues	
•	Sampling Strategy
•	Sample Location
•	Sample Number
•	Sample Matrix
•	Sample Volume/Mass
•	Discrete/Composite Samples
•	Sample Representativeness
•	Sampling Equipment, Containers and Preservatives
Analytical Issues	
•	Laboratory Sub-sampling
•	Sample Preparation Methods
•	Analytical Method
•	Detection Limits
•	Matrix Interferences
•	Bias
•	Holding Times
•	Calibration
•	Quality Control Results
•	Contamination
•	Reporting Requirements
•	Reagents/Supplies
Validation and Assessment	
•	Data Quality Objectives
•	Chain of Custody
•	Action Level
•	Completeness
•	Laboratory Audit Results
•	Field and Laboratory Records
•	Level of Uncertainty in Reported Values

5.3 Appropriate professionals must assess the project planning documents and completed project records to determine if the project findings match the conceptual model and decision logic. In areas where the findings don't match, the assessors must document and report their findings and, if possible, the potential impact on the decision process. Items subject to numerical confirmation are compared to the project plan and any discrepancies and their potential impact noted.

5.4 Effective quality control (QC) programs are those that empower the individuals performing the work to evaluate their performance and implement real-time corrections during the sampling or measurement process, or both. When quality control processes (including documentation) are properly implemented, they result in data sets (see Fig. 1) that are generated by in-control processes or out-of control processes that were not amenable to corrective action but whose details are explained by the project staff conducting the work. Good QC programs lead to reliable data that are seldom called into question during the assessment process. However, in cases where the absence of staff responsibility or authority to self-monitor and correct deficiencies at the working level is missing, the burden of assuring data integrity is placed on the
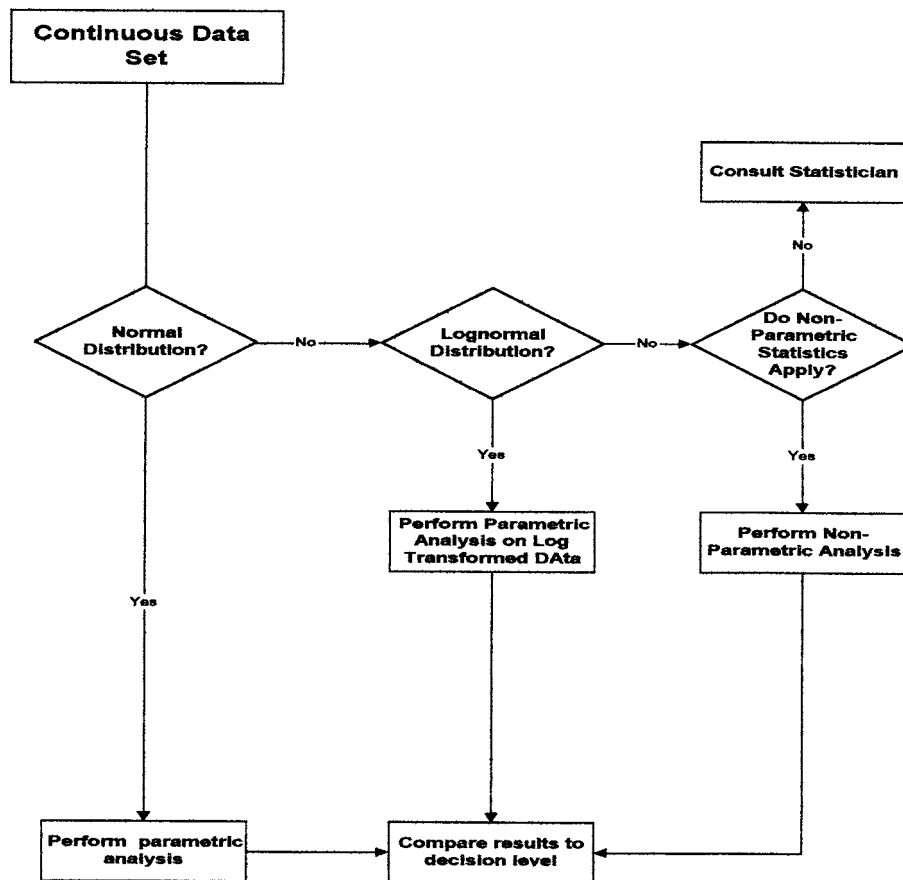
FIG. 1 General Strategy for Assessment of Continuous Data Sets

quality assurance (QA) function. The data assessment process must determine the location (working level or QA level) where effective quality control occurs (detection of error and execution of corrective action) in the data collection process and focus on how well the QC function was executed. As a general rule, if the QC function is not executed in real-time and thoroughly documented by the staff performing the work, the more likely the data assessor will be to find questionable data.

5.5 In addition to addressing the issues listed in Table 1, the data assessment process must search for unmeasurable factors whose impact cannot be detected by the review of the project records against expectations or numerical techniques. These are the types of things that are controlled by effective quality assurance programs, standard operating procedures, documentation practices, and staff training. Historically, efforts have been focused on the control of data collection errors through data review and the quality control process but little emphasis has been placed on the detection and evaluation of immeasur-

able errors using the quality assurance process. These unmeasurable sources of error are often the greatest source of uncertainty in the data collected for environmental projects. Examples of unmeasurable factors are given in Table 2.

5.6 Once the data assessment process has determined the degree to which the actual data collection effort met the expectations of the planners, the assessment process moves into the next phase to determine if the data generated by the effort can be verified and validated and whether it pass statistical tests for useability. These issues are discussed in the next sections.

6. Sources of Sampling Error

6.1 Sample collection may cause random or systematic errors. Random error affects the data by increasing the imprecision, whereas systemic error biases the data. The data assessment process should examine the available sampling records to determine if errors were introduced by improper sampling. A discussion of some of the more common sources of error follow.

6.1.1 *Random Error:*

6.1.1.1 Flaws in the sampling design which result in too few quality control samples being taken in the field can result in undetected errors in the sampling program. Adequate numbers of field QC samples (for example, field splits, co-located

TABLE 2 Examples of Unmeasurable Factors Affecting the Integrity of Environmental Data Collection Efforts

•	Biased Sampling/Subsampling	• Incorrect Dilutions
•	Sampling Wrong Area or Material	• Incorrect Documentation
•	Sample Switching (Mis-labeling)	• Matrix-Specific Artifacts
•	Misweighing/Misaliquoting	

samples, equipment rinse blanks, and trip blanks) are necessary to assess inconsistencies in sample collection practices, contaminated equipment, and contamination during the shipment process.

6.1.1.2 Variations (heterogeneity) in the media being sampled can cause concentration and property differences between and within samples. Field sampling and laboratory sub-sampling records should be examined to determine if heterogeneity was noted. This can explain wide variations in field and/or laboratory duplicate data.

6.1.1.3 Samples from the same population (including co-located samples) can be very different from each other. For example, one sample might be taken from a hot spot that was not visually obvious while the other was taken outside the perimeter of the hot spot. If data from areas of high concentration is contained in data sets consisting primarily of uncontaminated material, statistical outlier analysis might suggest the sample data should be omitted from consideration when evaluating a site. This can cause serious decision errors. Prior to declaring the data point(s) to be outliers, it is important for the assessor to examine the QC records from the analysis yielding the suspect data. If the QC data indicates the system was in control and review of the raw sample data reveals no handling or calculation errors, the suspect data should be discussed in the assessor's report but it should not be discounted. The site history and operating records may hold clues to the possible existence of hot spots.

6.2 *Systematic Error:*

6.2.1 Flaws in the sampling design that result in sampling of inappropriate locations can result in significant bias in the data. The samples collected from such a flawed plan will not be representative of the population and can result in incorrect decisions. The assessor should review the sampling plan for signs of potential bias and discuss their findings in the final report.

6.2.2 Sampling tools and equipment can deselect certain parts of a sample based on the physical properties (density, particle size, multi-phasic materials, particle geometry, etc.). If the sample is biased because of some physical characteristic, then any constituent that is distributed in the material based on that characteristic, will be incorrectly reported. Both field and laboratory sampling equipment can introduce this type of bias.

6.2.3 Incorrect sampling procedures can cause losses of certain constituents of a sample such as volatile organics. Failure to control the loss of of constituents that exist in the gaseous state often comprises the collection of unsaturated media for volatile compound characterization. Deterioration of the sample can also occur after collection due to improper storage and transportation. For example, samples left standing in sunlight or in a hot vehicle can undergo photochemical reactions or lose volatile constituents.

6.2.4 Interactions between the sample and the material of the sampling equipment or container, or both, are potential sources of positive or negative bias.

6.2.5 Inappropriate preservation of the sample can cause a shift in chemical equilibria, loss of target analytes, or degradation, or all of these. For example, when analyzing a water sample for dissolved metals, addition of nitric acid to a water sample containing suspended solids might dissolve metals from the solids, resulting in an incorrect high concentration being reported. Failure to preserve water samples intended for organic analysis may allow significant biological alteration of the sample.

6.2.6 The time of day and prevailing weather conditions when samples are collected can affect the sample. For example, strong winds can blow dust that can contaminate the samples. Cool mornings or evening can lead to higher retention of volatile components in near-surface soil samples compared to the samples collected in the heat of the day.

6.2.7 The above examples only serve to illustrate the need for an experienced professional to review the sampling activities and to place the resulting analytical data in the proper context of the sampling activity. Such assessments add materially to the usability of the data.

7. Sources of Analytical Error

7.1 Variation in the analytical process may cause random or systematic error. Random error affects the data by increasing the imprecision, whereas systematic error increases the bias of the data. The data assessment process should examine the available analytical records to determine if errors were introduced in the data by the analytical process. Analytical results can also be impacted by sample matrix effects. Discussion of some of the more common sources these types of error follow.

7.1.1 *Random Error:*

7.1.1.1 Random errors in the analytical process are often uncontrollable and unobserved. They are usually distributed between positive and negative error and tend to cancel out and so have little effect. However, for any one measurement, random error can be significant.

7.2 *Systematic Error*—The bias resulting from systematic error can be either positive or negative but it affects all results in a data set(s) the same way. Sources of systematic error are most often associated with sample preparation or analysis. Incomplete digestion or insufficient reaction time during sample preparation are examples that can produce negatively biased results during the preparation process. Improperly calibrated instruments, incorrect standards, dirty detectors, and leaking sample introduction systems are examples of instrumental problems that cause systematic error. They are most often detected when reference samples and laboratory control samples fail to produce the expected results.

7.3 *Sample Matrix Effects:*

7.3.1 The sample matrix can introduce either systematic or random error in analytical results. Consistently high or low results (systematic error) can be obtained when the matrix contains a non-target constituent that interferes with the accurate measurement of the target analyte. The interfering substance must be uniformly distributed in the matrix to produce consistent deviations from the true value. If the interference is non-uniformly distributed in the matrix, the error will appear as a random error.

7.3.2 The relationship between the sample matrix and the analytical method can result in an important class of matrix errors. When the method selected is not appropriate to the matrix, errors may result. One of the most common types of

mismatches of method and matrix is using methods designed for water analysis to analyze soils. Another is the use of methods designed for the analysis of naturally occurring materials, such as groundwater or soils, for the analysis of waste materials.

7.3.3 Most sample matrix and method selection errors can be detected by examining the results of matrix spike quality control samples where known amounts of the target analyte(s) are introduced into the sample before analysis. Spike results should be evaluated to determine the presence of any matrix effect. For certain types of analyses, simple dilution of the sample and re-analysis will demonstrate matrix effects when the second result, corrected for the dilution factor, is not consistent with the initial result.

8. Assessment of Environmental Data Sets

8.1 Data are usually verified and validated prior to comparing the results of environmental analysis to some decision level by suitable statistical processes. Data verification determines whether the laboratory carried out all steps required by the sampling and analysis plan or a contract, or both. After data is verified, it is validated. Validation examines the available laboratory data to determine whether an analyte is present or absent in a sample and the degree of overall uncertainty associated with the reported value. After data has been validated, it is normally compared to a decision level using suitable statistical techniques to determine the appropriate course of action.

8.2 The verification process compares the laboratory data package to a list of required data. These requirements are generated by two separate activities. The first is the contract for analytical services between the project and the laboratory and the second is the project sampling and analysis plan with its accompanying quality assurance project plan (QAPP) developed by project and laboratory staff. These two activities determine, *a priori*, the procedures the laboratory must use to produce data of known quality and the content of the analytical data package. Verification compares the material delivered by the laboratory against these requirements and produces a report that identifies those requirements which were not met (called *exceptions*). Verification exceptions normally identify:

8.2.1 Required steps not carried out by the laboratory (that is, incomplete analysis of all samples, lack of proper signatures, etc.),

8.2.2 Procedures not conducted at the required frequency (that is, too few blanks, duplicates, etc.),

8.2.3 Procedures which did not meet pre-set acceptance criteria (poor laboratory control sample recovery, unacceptable duplicate precision, etc).

8.3 The validation process begins with a review of the verification report or the laboratory data package, or both, to rapidly screen the areas of strength and weakness of the data set (tests of quality control). It continues with objective evaluation of sample data to confirm the presence or absence of an analyte (tests of detection) and to establish the statistical uncertainty (precision) of the measurement process for the analyte (test of uncertainty). Each data point is then qualified as to its integrity and dependability in the contest of all available laboratory data.

8.4 Examples of some important data project information that must be examined during the assessment of data are given in Table 1. Examples of some of the shortcomings that can occur are shown in Table 3. Some important characteristics of the data set that are frequently determined when examining quality control sample performance are given in Table 4. Data points not meeting the quality control criteria should be flagged and the magnitude and direction of any bias should be documented and made available for reference during the statistical evaluation processes that follow.

8.5 If project quality requirements are not met, further data assessment should not be undertaken until the data limitations are discussed with the project team. Data assessment cannot overcome basic design/execution flaws in the data collection process. Many times however, the project team can evaluate the problem and establish revised data quality objectives (different project expectations and new data requirements) factoring in the realities of the data collection effort which can then be used as the basis for data assessment.

TABLE 3 Common Data Requirements and Potential Shortcomings

Data Requirement	Potential Shortcomings
Number of samples	• Too few samples may have been collected or analyzed to be representative of the target population. • Too few samples were collected to narrow the estimate of the dispersion (variance, standard deviation, coefficient of variation, etc) of the measured results to acceptable levels.
Location of samples	• Samples were collected from the wrong locations due to error or inaccessibility.
Analyte/method	• Incorrect choice of analyte/method for the sample matrix
Quality control	• Measurement system not calibrated • Contamination found in field, trip, or method blanks • Method performance on reference samples unsatisfactory • Calculation errors
Method sensitivity	• Failure to meet minimum detectable limits
Method precision	• Failure to achieve satisfactory duplicate results for analysis of field samples due to sample characteristics or other analytical problems
Method bias	• Failure to demonstrate method performance on reference materials or analytical standards • Failure to demonstrate satisfactory target analyte spike/surrogate recoveries in field sample analysis
Interferences	• Presence of unanticipated materials/analytes in field samples that render accurate analysis suspect
Action level	• Not provided

TABLE 4 Information Derived From Quality Control Samples[A]

Type of QC Sample	Type of Information											
	Precision			Bias			Contamination					
	Sampling	Splitting	Preparation and Analysis	Spiking	Field/ Shipping/ Storage	Laboratory	Containers and Preservatives	Field Environment	Equipment	Cross-Contamination	Laboratory	
Replicates												
Splits, field		X	X									
Collocated, field	X		X									
Splits, laboratory		X	X									
Spikes												
Field				X	X	X						
Laboratory, matrix				X		X						
Blanks												
Trip							X			X		
Field							X	X		X		
Equipment							X	X	X	X		
Method											X	

[A] Can be assessed using numerical techniques.

9. Statistical Evaluation of Data Sets

9.1 The US EPA *Guidance for Data Quality Assessment, QA/G-9* (**1**)[3] is a good source for information on the following statistical approaches to data assessment.

9.1.1 *Continuous Data:*

9.1.1.1 Continuous data are data where the values of the individual samples may vary form zero to any maximum value. Examples of continuous data are the concentration of a constituent in soil or the percent moisture in an environmental sample. This is the type of information most frequently collected in environmental waste management projects. It is normally used to establish a statistical characteristic of the target population which is then compared to a decision level resulting in an action. This is referred to as the "decision rule" and normally takes the form:

If (characteristic of the population) (method of comparison) (action level), then (action). Otherwise, (alternate action).

where the items in parentheses are determined by the project team on a project-specific basis. Two examples are:

If (the average concentration of mercury in the top 15 cm of soil over the site) (is greater than) (100 mg/kg), then (excavate the top 30 cm of soil and dispose of in a RCRA landfill). Otherwise, (no remediation is required).

and:

If (less than one half the randomly selected waste oil drums have an average organic halide concentration) (of less than 500 ppm), then (composite the contents of all drums and use it as boiler fuel). Otherwise, (send all drums to a RCRA treatment and disposal facility).

9.1.2 Before beginning the statistical interpretation of a continuous data set, plots of the data should be constructed to guide the statistical interpretation of the data that follows. Examples of the types of plots that can be constructed are:

9.1.2.1 Concentration versus time, and

9.1.2.2 Concentration versus location in two or three dimensions as appropriate

9.1.2.3 These types of plots provide a picture of the distribution of the parameter of interest and permit the identification of strata as a function of time or location. Plots also identify data points which are abnormally high or low with respect to the surrounding data. These are potential outliers and they can be more rigorously evaluated by the verification and validation process to determine whether there is an analytically-related explanation. This information will identify random or stratified data sets and outliers or QC-failed data prior to statistical evaluation.

9.1.3 *Normally Distributed Data:*

9.1.3.1 Once the data evaluation described above have been completed, statistical techniques should be used to evaluate the data against the decision criteria. The key steps in the sequence to evaluate continuous data are shown in Fig. 2.

9.1.3.2 The first step is to determine if the data are normally distributed. That is, are there an approximately equal number of values that are less than and greater than the mean and is the range of values approximately equal on either side of the mean (See Fig. 2). This property of normal distribution is a reasonable model of the behavior of certain random phenomena and can be used to approximate many kinds of data.

9.1.3.3 There are several graphical techniques that can be applied to determine if data are normally distributed. Among them are: stem- and leaf- diagrams, histogram/frequency plots, box and whiskers plots, ranked data plots, quantile plots, and, normal probability plots (quantile-quantile plots).

9.1.3.4 The use of plots to determine if data are normally distributed involves a subjective decision on the part of the

---

[3] The boldface numbers given in parentheses refer to a list of references at the end of the text.
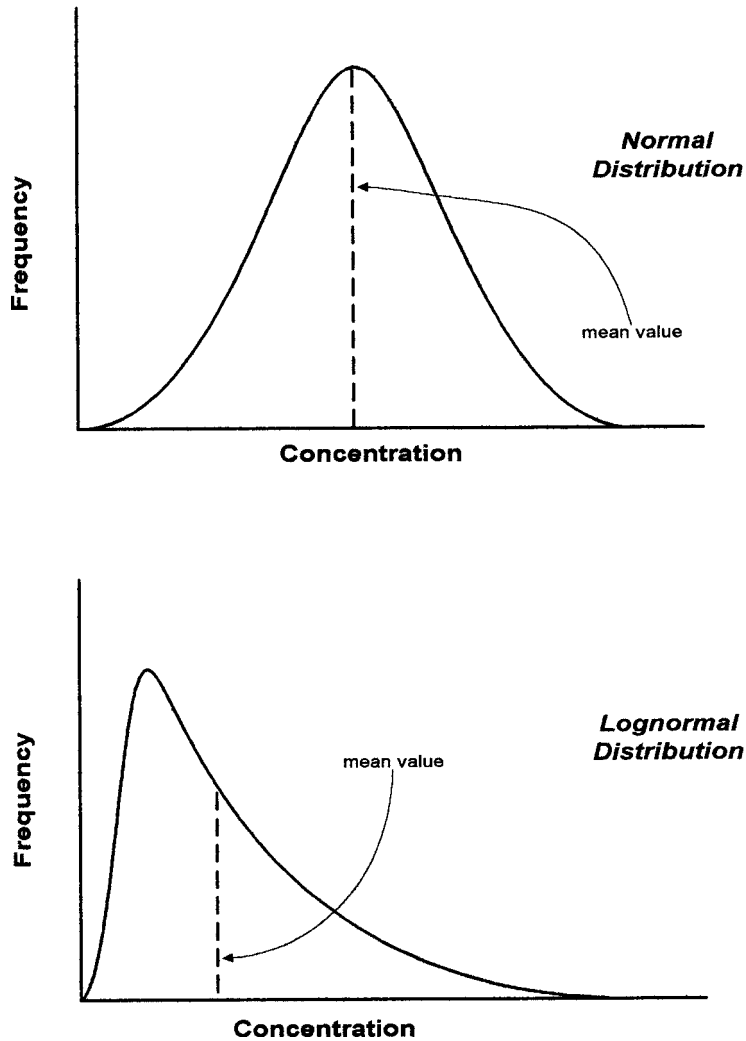
FIG. 2 Two Types of Data Distribution

individuals making the assessment. This is easy when the data are very non-normal but more difficult as the data approach normal distribution. There are series of formal numerical methods to test for normal distribution. The Shapiro - Wilkes test can be applied to data sets of less than 50 samples. For larger size sample sets (up to 1000 data points), Fillben's Statistic is frequently used. Both methods are difficult to implement by hand because of the large number of calculations required but are readily accomplished by computer programs.

9.1.4 Once the normal distribution of the data is shown, the straightforward calculation of the statistical quantities used in the project decision rule can be performed. For example, the two-sided confidence limits for the mean (that is a parametric population characteristic) can be performed. This allows the data user to determine the interval in which the true mean is expected to be found with specified confidence. The mean lead level, interval and confidence are frequently expressed as:

the level of lead in the soil is X ± x at the 95 % confidence level

9.1.4.1 The width of the interval, ± 2x, can be calculated for varying degrees of confidence (selected by the data user) to

meet project-specific tolerable error rates for making the correct decision. For a two-sided confidence interval, x is given by:

$$x = \frac{\left(t_{97.5,\,n-1}\right)(s)}{\sqrt{n}} \qquad (1)$$

where:

s = standard deviation,
n = the number of samples, and
t = the student t-statistic.

For the one-sided confidence interval, x is given by:

$$x = \frac{\left(t_{0.95,\,n-1}\right)(s)}{\sqrt{n}} \qquad (2)$$

9.1.5 Some types of statistical quantities which can be calculated from normally distributed data include, but are not limited to: mean, range, variance, standard deviation, coefficient of variation, and, confidence limits.

9.1.6 The choice of which statistics should be calculated is dependent on the characteristic of the population that will be

7

used in the decision rule. After the appropriate statistical quantities are calculated from the field sample data, they should be compared to the assumed values which were the basis of the DQO calculation of tolerable error rate at the decision level.

9.1.6.1 Fig. 3 shows four examples of the comparison of various types of laboratory data against a regulatory decision level. In the examples, the upper confidence limit of the data set is compared to regulatory decision level. The figure represents four common characteristics of continuous data sets.

These are: unbiased and precise, unbiased and imprecise, biased and precise, biased and imprecise.

9.1.6.2 To distinguish between the biased and unbiased situations pictured in the figure, one can refer to the verification and validation results described previously. The positive bias displayed in the bottom two examples of Fig. 3 should be reflected in high matrix spike recoveries and higher than normal recoveries on reference samples or laboratory control samples. None of the examples in Fig. 3 reflect sampling bias, they only apply to analytical bias.
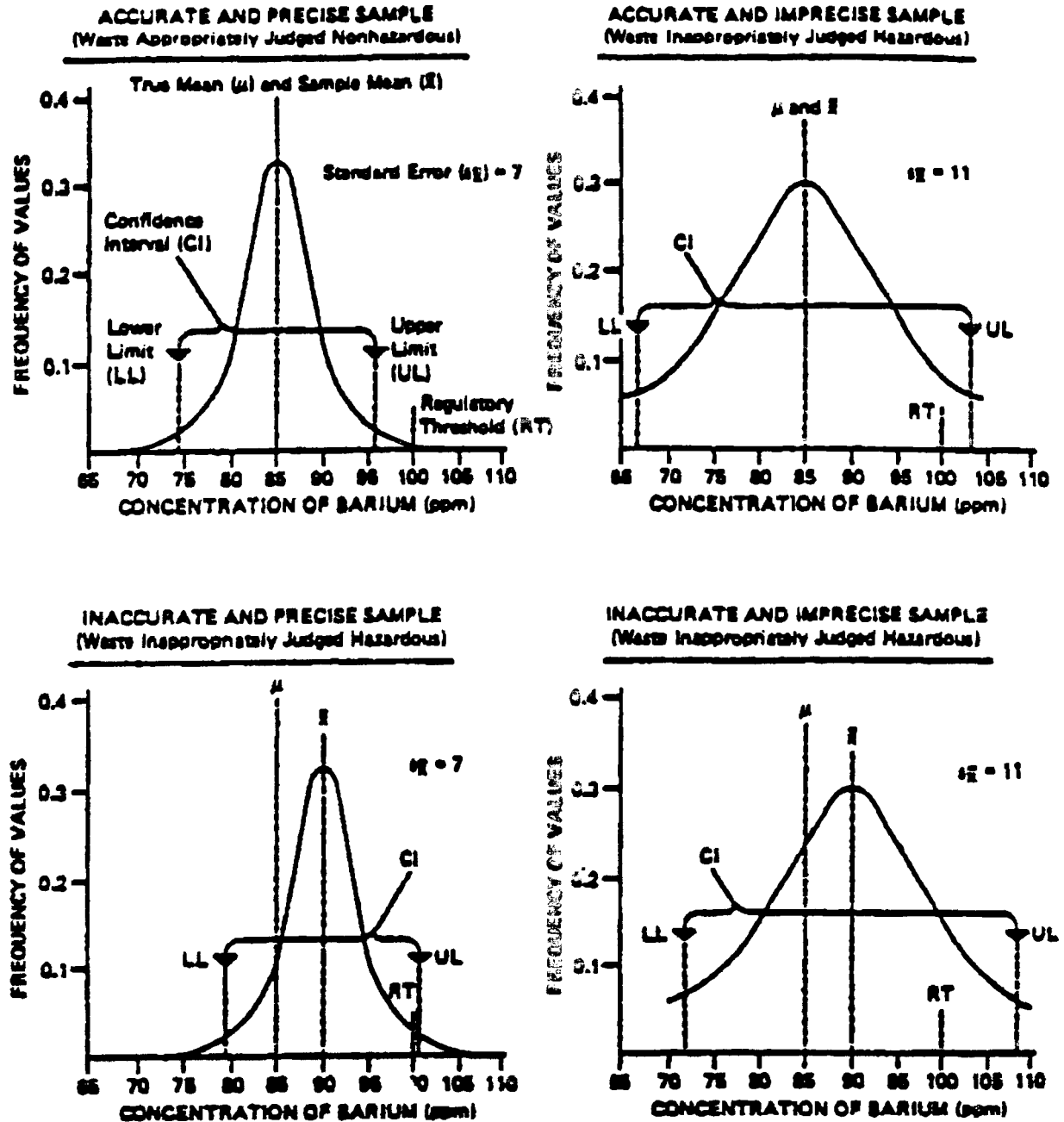
FIG. 3 Four Examples of Laboratory Data

9.1.6.3 The results shown in Fig. 3 may also prompt a re-examination of the decision criteria reached in the DQO process. The impact of imprecision and bias on the decision making process increases as the mean approaches the regulatory threshold (or action level) and as imprecision increases. At some mean value, increased variance in the mean will determine that a correct decision cannot be reached with acceptable certainty within the project budget.

9.1.6.4 The opposite of the above example may also occur, the variance may be much less than estimated and the mean value of the parameter of interest may be low compared to the action level. It follows then, that the decision error rate and overall certainty will be improved so that a much higher degree of confidence in the final decision can be attained.

9.1.7 The complex inter-relationship between confidence level, relative error, action levels, variance, number of samples, and population characteristic will determine if an acceptable decision can be reached. Therefore, the data assessor must evaluate if the final outcome meets the data quality objective level of confidence and is within the tolerable error window. If so, the data can be said to be acceptable for making the intended decision. If not, the DQO objectives for the final decision must be changed or additional sampling and analysis must be conducted to meet the objectives. Table 5 (2) shows how the two variables, level of error and variance, impact the correct number of samples.

9.1.7.1 The assessor must determine whether the data conform to the project design criteria for level of error and coefficient of variation in the data for the desired confidence level in the decision when n samples were collected and analyzed. If the design criteria were not met, the decision makers can take additional samples, use more precise analytical methods, or accept a lower confidence level.

9.1.8 *Non-Normally Distributed Data:*

9.1.8.1 Not all environmental decisions are based on normally distributed data. When the data are not normal, non-parametric statistical methods can sometimes be used. An example is the use of non-parametric tolerance limits(3). Suppose a compliance limit limit of 25 mg/kg of copper is in place at a soil remediation project. Further suppose that after the collection of six samples, non of them exceeded this limit. The values were 9, 11, 13, 15, 18, and 24 mg/kg. The question may be asked whether a total of six samples is large enough to allow for the making of a complinace decision with a high level of confidence. In the context of non-parametric tolerance limits the question can be expressed as:

What is the probability of of exceeding the largest measured value of 24 mg/kg at some level of confidence (for example, 95 %) when there are only six sample values?

9.1.8.2 When there is confidence that this probability is small, then a conclusion of compliance can be made. Conversely, if the probability exceeds the limits set by the stakeholders then the question of how many more samples not exceeding 24 mg/kg are needed to make the conclusion of compliance. The determination of the necessary number of samples to reach a given level of confidence is provided in Table 6 (2).

9.1.8.3 It can be seen from the table that, given a total of six samples, the it is 95 % probable that at least 50 % of the values from this site do not exceed 24 mg/kg. If the stakeholders prefer an 80 % degree of confidence, then a total of fourteen samples not exceeding 24/kg are needed.

9.1.9 Many environmental data sets represents populations where the parameter of interest approaches a lower theoretical limit (that is, zero concentration of contaminant in soil or water). Such data sets are not normally distributed, and most values approach zero with a decreasing number of values as the concentration increases. The probability model that most often describes these properties is the lognormal distribution. A graph of this distribution is shown in Fig. 2.

9.1.9.1 The project design may address this requirement in one of two ways. Composite samples can be collected and analyzed rather than a series of individual or discrete samples. The process of compositing physically averages out the higher valued samples with the much larger number of lower valued samples.

9.1.9.2 In one commonly used approached, the data set can be transformed (changed by the application of a mathematical process to each data point) into a normally distributed data set by taking the natural logarithm of the data.

9.1.9.3 Population data sets that have been normalized by either composite sampling or logarithmic transformation are then treated as normally distributed data. This means the same statistical reductions can be used to yield the population characteristic (for example, mean, variance, upper and lower confidence limits) used in the decision rule and the same hypothesis test to determine the correct action. It is important that whenever using transformed data to determine the statistical parameter used in the decision rule, the value of the action level (or regulatory threshold) must be transformed as well. It is also acceptable to take the antilog of the calculated population statistic and compare that to the action level.

9.1.9.4 If a suitable transformation of the population data base cannot be found which results in a normal distribution, more advanced statistical technique may be required (see

TABLE 5 Number of Samples as a Function of the Coefficient of Variation and Level of Error at the 95 % Confidence Limit

	Number of Samples				
	Coefficient of Variation				
	0.1	0.5	0.65	1	2
Relative Error					
0.1	4	96	162	384	1537
0.2	1	24	40	96	384
0.3	. . .	10	18	42	170
0.5	. . .	4	6	15	61
1.0	. . .	1	2	4	15
2.0	. . .	. . .	. . .	1	4

TABLE 6 Relationship of Degree of Confidence to the Percentage of the Population< Maximum Measured Value of 24 (n = 6)

Degree of Confidence	Percentage of Population < Maximum Measured Value
98	50
85	70
80	75
70	80

statistical texts and monographs). The project team can also be asked to restate the decision rule using population characteristics that don't require a normally distributed data set (non-parametric characteristics).

10. Evaluation of Discrete Data

10.1 Discrete data are made up of a series of sample results that are expressed as a simple pass/fail, yes/no, or positive/ negative. This is frequently described as *dichotomous data*. Examples of analytical test that generate dichotomous data are flash point and corrosivity.

10.2 In most cases, each individual data point represents a target population (that is, the contents of a drum) and the decision is made by comparing the individual value with the action level. For example, each drum sample that ignites when tested by a flammability test is determined to have failed the test (the dichotomous response) and the project response is to send the drum to a hazardous waste treatment and disposal facility. In this case (each sample represents the target population) no statistical assessment of the data is possible, and the decision is made on a qualitative basis. In most cases, the test results supported by the data requirements in Table 3 become the foundation for the qualitative determination.

10.3 In cases where a set of discrete samples from a large population (that is, 25 individual drum samples randomly selected from a set of 125 drums) are analyzed by a test producing dichotomous results, the proportion of drums failing the test is the statistic used in the decision rule. Tests of proportion require that the data set be normally distributed or capable of being transformed into a normally distributed set.

11. Outliers

11.1 Individual measurements that are extremely large or small relative to the majority of the data are generally suspected of misrepresenting the target population from which they were collected. These values are generally referred to a outliers. Outliers may be the result of sampling or analytical errors, or both, or may represent true extremes in the population. Before determining the value of population statistic used in the decision rule, statistical outlier tests should be performed to identify outliers which are then carefully investigated to see if the value may be the result of an error.

11.1.1 There are four commonly used techniques to evaluate data for outliers.

11.1.1.1 *Comparison to Historical Data*—If sufficient data exist over time from the same site or population, potential outlying values can be compared to past data from the same sampling point or location. If past data show the same high or low values as the current data, it is reasonable to assume that the current extreme values are valid. The data assessor should describe their findings in the evaluation report and a description of how the extreme values were used in past decision making processes provided. This will allow project decision makers to apply consistent decision logic over time.

11.1.1.2 *Trend Analysis*—As discussed in 9.1.2, plots of data points versus time or three dimensional location will reveal trends in data sets. If a trend exist, it can be used to support the validity of data points at the high or low extremes. Data points

that are inconsistent with the trend should be more carefully evaluated to determine if they are real.

11.1.1.3 *Comparison with Companion Data Sets*—The existence of a correlationship between two data sets from a single population (for example, total petroleum hydrocarbons and volatile organic compounds in soil) is strong support for the validity of extreme values. If samples show both high total petroleum hydrocarbon (TPH) and toluene values, then both data points can be considered valid. A clear picture of this type of data relationship can be gained by plotting the two values for a single sample (TPH and toluene) on a set of coordinates keyed to two variables. The points that trend in the same general straight line can be considered valid. Points that lie off the line may do so because one or both of the variables is an outlier.

11.1.1.4 *Checks for Errors*—Errors in interpreting sampling instructions (location, sampling equipment, sample preservation), errors in analysis (incorrect sample size, uncalibrated instruments), errors in calculations or reporting can lead to apparent outliers in 11.1.1.1 through 11.1.1.3 above. The data assessor should follow up on potential outliers identified in the above processes and check for errors. If errors can be found and documented, the corrected data points should be introduced to the data set and the entire set re-evaluated.

11.2 Suspect values that can be documented based on some scientific observation or quality assurance basis should be omitted from the calculation of the statistic. They are reported to the data user in the final report but they are flagged as outlying values not to be used in the project decision or calculation of the population statistic (for example, mean value). If no error can be found, the outlying value should be retained but its presence and impact on the decision statistic clearly reported to the decision maker who may ask for re-sampling and analysis in critical situations.

11.3 There are other more powerful statistical tools that can be applied to the detection and treatment of outliers but they should be applied under the guidance of an experienced statistician.

12. Non-Detect Values

12.1 The treatment of non-detect values in a population data base can greatly affect the final population statistic. There are a variety of ways to treat values that lie below the detection limit of the analytical method. However, there are no general procedures that are applicable in all cases. The choice of data analysis method is dependent on the percentage of samples in the data base that are below the detection limit. The EPA's *Guidance for Data Quality Assessment* (1) provides a general discussion of several approaches outlined below. A brief discussion of some generally accepted approaches follows.

12.2 *Less than 15 % Non-Detects*—When less than 15 % of the reported data falls below the detection limit, it is possible to replace the non-detected values with a small number less than the detection limit. The number most frequently chosen is one-half the detection limit. For difficult to analyze matrices, one-half the practical quantitation limit can be used if the PQL has been determined for the analyses in the matrix being analyzed.

12.3 *Between 15 to 50 % Non-Detects*—There are three methods that can be applied when 15 to 50 % of the data are non-detects.

12.3.1 Cohen's method uses the sample mean, the variance, the total number of observations below the detection limit, and a set of statistical tables to determine corrected values for the sample mean and variance. Cohen's method is constrained by the requirement that the data above the detection limit be normally distributed and the detection limit always be the same.

12.3.2 The method of trimmed means determines the number of observations below the detection limit and directs that this number of measurements be deleted from both the high end and low end of the data. The remaining data is used to calculate the trimmed population mean and variance.

12.3.3 The method of winsorized means replaces the set non-detects (*n*-observations) with values equal to the *n*-lowest values above the detection limit. The highest values are adjusted downward in a similar fashion, that is, the *n*-highest values are dropped and replaced by the series of *n*-highest values remaining in the data set.

12.4 At the interval between 50 to 90 % non-detects, a test of proportions is recommended. For sets where less than 10 % of the values are above the detection limit, a Poisson Distribution approach is recommended.

12.5 All the methods mentioned here can be found in more detail in standard statistical texts.

13. Final Use of Data in the Decision Rule

13.1 Once it has been established that the data meet the assumptions of the DQO process in terms of anticipated values using this assessment process, they are tested against the action level given in the decision rule using hypothesis tests. These tests determine if the data meet or exceed the action level and thereby direct the data user to the agreed upon course of action for the project. A discussion of the types of hypothesis testing is beyond the scope of this standard. Guidance can be found in standard statistical texts and monologues

REFERENCES

(1) Environmental Protection Agency, *Guidance for Data Quality Assurance, EPA QA/G-9*, 1996.
(2) Gilbert, R. O., *Statistical Methods for Environmental Pollution Monitoring*, Van Nostrand Reinhold, New York, 1987.
(3) Owen, D.B., 1962, *Handbook of Statistical Tables*, Addison-Wesley Publishing Company, Reading, Massachusetts.
(4) Cochran, W.G., *Sampling Techniques*, 3rd Ed., John Wiley and Sons, Inc., New York, 1977.
(5) Desu, M.M. and D. Raghavarao, *Sample Size Methodology*, Academic Press, San Diego, 1990.