United States Environmental Protection Agency | Solid Waste and Emergency Response (5305W) | EPA530-D-02-002 August 2002 www.epa.gov/osw

Office of Solid Waste

RCRA Waste Sampling Draft Technical Guidance

Planning, Implementation, and Assessment

EPA530-D-02-002
August 2002

RCRA Waste Sampling Draft Technical Guidance

Planning, Implementation, and Assessment

Office of Solid Waste
U.S. Environmental Protection Agency
Washington, DC 20460

DISCLAIMER

The United States Environmental Protection Agency's Office of Solid Waste (EPA or the Agency) has prepared this draft document to provide guidance to project planners, field personnel, data users, and other interested parties regarding sampling for the evaluation of solid waste under the Resource Conservation and Recovery Act (RCRA).

EPA does not make any warranty or representation, expressed or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report. EPA does not assume any liability with respect to the use of, or for damages resulting from the use of, any information, apparatus, method, or process disclosed in this report. Reference to trade names or specific commercial products, commodities, or services in this report does not represent or constitute an endorsement, recommendation, or favoring by EPA of the specific commercial product, commodity, or service. In addition, the policies set out in this document are not final Agency action, but are intended solely as guidance. They are not intended, nor can they be relied upon, to create any rights enforceable by any party in litigation with the United States. EPA officials may decide to follow the guidance provided in this document, or to act at variance with the guidance, based on an analysis of specific site or facility circumstances. The Agency also reserves the right to change this guidance at any time without public notice.

ACKNOWLEDGMENTS

Development of this document was funded, wholly or in part, by the United States Environmental Protection Agency (U.S. EPA) under Contract No. 68-W6-0068 and 68-W-00-122. It has been reviewed by EPA and approved for publication. It was developed under the direction of Mr. Oliver M. Fordham, Office of Solid Waste (OSW) and Kim Kirkland (OSW) in collaboration with Dr. Brian A. Schumacher, Office of Research and Development (ORD). This document was prepared by Mr. Robert B. Stewart, Science Applications International Corporation (SAIC). Additional writers included Dr. Kirk Cameron (MacStat Consulting, Ltd.), Dr. Larry P. Jackson (Environmental Quality Management), Dr. John Maney (Environmental Measurements Assessment Co.), Ms. Jennifer Bramlett (SAIC), and Mr. Oliver M. Fordham (U.S. EPA).

EPA gratefully acknowledges the contributions of the technical reviewers involved in this effort, including the following:

U.S. EPA Program Offices

Deana Crumbling, TIO
Evan Englund, ORD
George Flatman, ORD
Joan Fisk, OERR
David Friedman, ORD
Chris Gaines, OW
Gail Hansen, OSW
Barnes Johnson, OSW
Joe Lowry, NEIC
John Nocerino, ORD
Brian A. Schumacher, ORD
Jim Thompson, OECA
Jeff Van Ee, ORD
Brad Venner, NEIC
John Warren, OEI

U.S. EPA Regions

Dan Granz, Region I
Bill Cosgrove, Region IV
Mike Neill, Region IV
Judy Sophianopoulos, Region IV
Brian Freeman, Region V
Gene Keepper, Region VI
Gregory Lyssy, Region VI
Bill Gallagher, Region VI
Deanna Lacy, Region VI
Maria Martinez, Region VI
Walt Helmick, Region VI
Charles Ritchey, Region VI
Terry Sykes, Region VI
Stephanie Doolan, Region VII
Dedriel Newsome, Region VII
Tina Diebold, Region VIII
Mike Gansecki, Region VIII
Roberta Hedeen, Region X
Mary Queitzsch, Region X

ASTM Subcommittee D-34

Brian M. Anderson, SCA Services
Eric Chai, Shell
Alan B. Crockett, INEL
Jim Frampton, CA DTSC
Susan Gagner, LLNL
Alan Hewitt, CRREL
Larry Jackson, EQM
John Maney, EMA

Other Organizations

Jeffrey Farrar, U.S. Bureau of Reclamation
Jeff Myers, Westinghouse SMS
Rock Vitale, Environmental Standards
Ann Strahl, Texas NRCC

CONTENTS

LIST OF ACRONYMS

AL	Action Level
ASTM	American Society for Testing and Materials
BDAT	Best Demonstrated Available Technology
BIF	Boiler and Industrial Furnace
CERCLA	Comprehensive, Environmental Response, Compensation & Liability Act
CFR	Code of Federal Regulations
DOT	Department of Transportation
DQA	Data Quality Assessment
DQO	Data Quality Objective
EA	Exposure area
FR	Federal Register
HWIR	Hazardous Waste Identification Rule (waste)
IATA	International Air Transport Association
ICR	Ignitability, Corrosivity, and Reactivity
IDW	Investigation-derived waste
LCL	Lower confidence limit
LDR	Land Disposal Restrictions
ORD	Office of Research and Development
OSHA	Occupational Safety and Health Administration
OSW	Office of Solid Waste
PBMS	Performance-based measurement system
ppm	Parts per million
QAD	Quality Assurance Division
QAPP	Quality Assurance Project Plan
QA/QC	Quality Assurance/Quality Control
RCRA	Resource Conservation and Recovery Act
RT	Regulatory Threshold
SOP	Standard operating procedure
SWMU	Solid waste management unit
TC	Toxicity Characteristic
TCLP	Toxicity Characteristic Leaching Procedure
TSDF	Treatment, storage, or disposal facility
UCL	Upper confidence limit
USEPA	U.S. Environmental Protection Agency (we, us, our, EPA, the Agency)
UTS	Universal Treatment Standard
VOC	Volatile organic compound
WAP	Waste analysis plan

RCRA WASTE SAMPLING
DRAFT TECHNICAL GUIDANCE

1 INTRODUCTION

1.1 What Will I Find in This Guidance Document?

You'll find recommended procedures for sampling solid waste under the Resource Conservation and Recovery Act (RCRA). The regulated and regulatory communities can use this guidance to develop sampling plans to determine if (1) a solid waste exhibits any of the characteristics of a hazardous waste[1], (2) a hazardous waste is prohibited from land disposal, and (3) a numeric treatment standard has been met. You also can use information in this document along with that found in other guidance documents to meet other sampling objectives such as site characterization under the RCRA corrective action program.

This guidance document steps you through the three phases of the sampling and analysis process shown in Figure 1: planning, implementation, and assessment. Planning involves "asking the right questions." Using a systematic planning process such as the Data Quality Objectives (DQO) Process helps you do so. DQOs are the specifications you need to develop a plan for your project such as a quality assurance project plan (QAPP) or a waste analysis plan (WAP). Implementation involves using the field sampling procedures and analytical methods specified in the plan and taking measures to control error that might be introduced along the way. Assessment is the final stage in which you evaluate the results of the study in terms of the original objectives and make decisions regarding management or treatment of the waste.

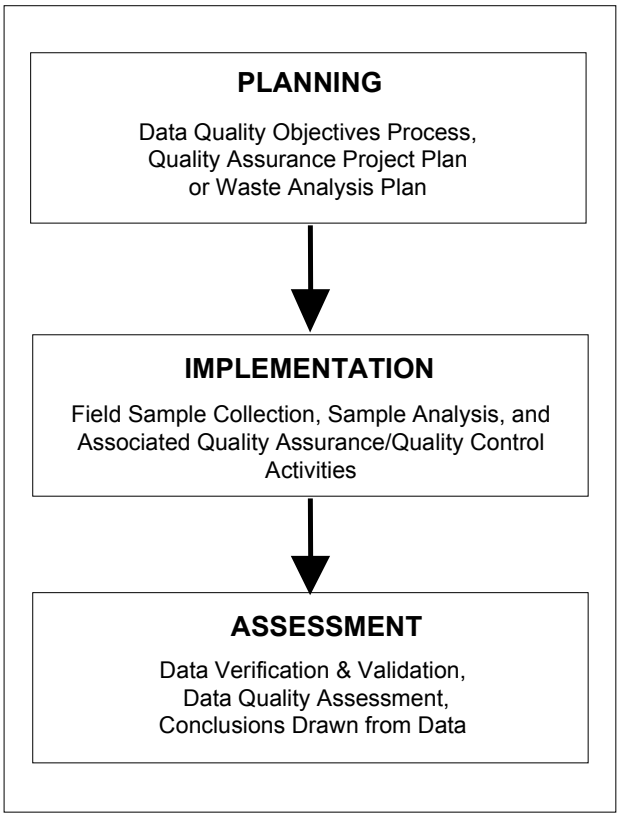

Figure 1. QA Planning and the Data Life Cycle (after USEPA 1998a).

1.2 Who Can Use This Guidance Document?

Any person who generates, treats, stores, or disposes of solid and hazardous waste and conducts sampling and analysis under RCRA can use the information in this guidance document.

[1] If a solid waste is not excluded from regulation under 40 CFR 261, then a generator must determine whether the waste exhibits any of the characteristics of hazardous waste. A generator may determine if a waste exhibits a characteristic either by testing the waste or applying knowledge of the waste, the raw materials, and the processes used in its generation.

For the development of a technically sound sampling and project plan, seek competent advice during the initial stages of project design. This is particularly true in the early developmental stages of a sampling plan when planners need to understand basic statistical concepts, how to establish objectives, and how the results of the project will be evaluated.

This document is a practical guide, and many examples are included throughout the text to demonstrate how to apply the guidance. In addition, we have included a comprehensive glossary of terms in Appendix A to help you with any unfamiliar terminology. We encourage you to review other documents referenced in the text, especially those related to the areas of sampling theory and practice and the statistical analysis of environmental data.

1.3 Does This Guidance Document Replace Other Guidance?

EPA prepared this guidance document to update technical information contained in other sources of EPA guidance such as Chapter Nine "Sampling Plan" found in *Test Methods for Evaluating Solid Waste, Physical/Chemical Methods,* EPA publication *SW-846* (1986a). This draft guidance document does not replace SW-846 Chapter Nine, nor does it create, amend, or otherwise alter any regulation. Since publication of SW-846 Chapter Nine, EPA has published a substantial body of additional sampling and statistical guidance documents that support waste and site characterization under both RCRA and the Comprehensive, Environmental Response, Compensation & Liability Act (CERCLA) or "Superfund." Most of these guidance documents, which focus on specific Agency regulations or program initiatives, should continue to be used, as appropriate. Relevant EPA guidance documents, other references, and resources are identified in Appendix B and throughout this document.

In addition to RCRA program-specific guidance documents issued by EPA's Office of Solid Waste (OSW), EPA's Office of Environmental Information's Quality Staff has developed policy for quality assurance, guidance documents and software tools, and provides training and outreach. For example, the Quality Staff have issued guidance on the following key topic areas:

- The data quality objectives process (USEPA 2000a, 2000b, and 2001a)
- Preparation of quality assurance project plans (USEPA 1998a and 2001b) and sampling plans (2000c)
- Verification and validation of environmental data (USEPA 2001c)
- Data quality assessment (USEPA 2000d).

Information about EPA's Quality System and QA procedures and policies can be found on the World Wide Web at http://www.epa.gov/quality/.

If you require additional information, you should review these documents and others cited in this document. In the future, EPA may issue additional supplemental guidance supporting other regulatory initiatives.

Finally, other organizations including EPA Regions, States, the American Society for Testing and Materials (ASTM), the Department of Defense (e.g., the Air Force Center for Environmental

Excellence), and the Department of Energy have developed a wide range of relevant guidance and methods. Consult these resources for further assistance, as necessary.

1.4 How Is This Document Organized?

As previously indicated in Figure 1, this guidance document covers the three components of a sampling and analysis program: planning, implementation, and assessment. Even though the process is pictured in a linear format, in practice a sampling program should include feedback between the various components. You should review and analyze data as collected so you can determine whether the data satisfy the objectives of the study and if the approach or objectives need to be revised or refined, and so you can make reasoned and intelligent decisions.

The remaining sections of this guidance document address specific topics pertaining to various components of a sampling program. These sections include the following:

Section 2 - Summary of RCRA Regulatory Drivers for Waste Sampling and Analysis – This section identifies and summarizes the major RCRA programs that specify some sort of sampling and testing to determine if a waste is a hazardous waste, to determine if a hazardous waste treatment standard is attained, and other determinations.

Section 3 - Fundamental Statistical Concepts -- This section provides an overview of fundamental statistical concepts and how the sample analysis results can be used to classify a waste or determine its status under RCRA. The section serves as a refresher to those familiar with basic statistics. In those cases where you require more advanced techniques, seek the assistance of a professional environmental statistician. Detailed guidance on the selection and use of statistical methods is provided in Section 8 and Appendix F.

Section 4 - Planning Your Project Using the DQO Process -- The first phase of sampling involves development of DQOs using the DQO Process or a similar structured systematic planning process. The DQOs provide statements about the expectations and requirements of the data user (such as the decision maker).

Section 5 - Optimizing the Design for Obtaining the Data -- This section describes how to link the results of the DQO Process with the development of the QAPP. You optimize the sampling design to control sampling errors within acceptable limits and minimize costs while continuing to meet the sampling objectives. You document the output of the DQO Process in a QAPP, WAP, or similar planning document. Here is where you translate the data requirements into measurement performance specifications and QA/QC procedures.

Section 6 - Controlling Variability and Bias in Sampling -- In this section, we recognize that random variability and bias (collectively known as “error”) in sampling account for a significant portion of the total error in the sampling and analysis process – far outweighing typical analytical error. To address this concern, the section describes the sources of error in sampling and offers some strategies for minimizing those errors.

Section 7 - Implementation: Selecting Equipment and Conducting Sampling -- In this section, we describe the steps for selecting sampling equipment based on the physical and chemical characteristics of the media to be sampled and the type of RCRA unit or location from which the samples will be obtained. The section provides guidance on field sampling activities, such as documentation, chain-of-custody procedures, decontamination, and sample packaging and shipping. Finally, guidance is provided on sample homogenization (or mixing), splitting, and subsampling.

Section 8 - Assessment: Analyzing and Interpreting Data -- Once you have obtained the data in accordance with the elements of the QAPP or WAP, you should evaluate the data to determine whether you have satisfied the DQOs. Section 8 describes the data quality assessment (DQA) process and the statistical analysis of waste-sampling data.

Appendix A - Glossary of Terms -- This appendix comprises a glossary of terms that are used in this document.

Appendix B - Summary of RCRA Regulatory Drivers for Conducting Waste Sampling and Analysis -- An overview of the RCRA regulatory requirements and other citations related to waste sampling and testing is provided in this appendix.

Appendix C - Strategies for Sampling Heterogeneous Wastes -- The heterogeneity of a waste or media plays an important role in how you collect and handle samples and what type of sampling design you use. This appendix provides a supplemental discussion of large-scale heterogeneity of waste and its impact on waste-sampling strategies. Various types of large-scale heterogeneity are identified and techniques are described for stratifying a waste stream based on heterogeneity. Stratified sampling can be a cost-effective approach for sampling and analysis of heterogeneous wastes.

Appendix D - A Quantitative Approach for Controlling Fundamental Error -- The mass of a sample can influence our ability to obtain reproducible analytical results. This appendix provides an approach for determining the appropriate mass of a sample of particulate material using information about the size and shape of the particles.

Appendix E - Sampling Devices -- This appendix provides descriptions of recommended sampling devices. For each type of sampling device, information is provided in a uniform format that includes a brief description of the device and its use, advantages and limitations of the device, and a figure to indicate the general design of the device. Each summary also identifies sources of other guidance on each device, particularly any relevant ASTM standards.

Appendix F - Statistical Methods -- This appendix provides statistical guidance for the analysis of data generated in support of a waste-testing program under RCRA.

Appendix G - Statistical Tables -- A series of statistical tables needed to perform the statistical tests used in this guidance document are presented here.

Appendix H - Statistical Software -- A list of statistical software and “freeware” (no-cost software) that you might find useful in implementing the statistical methods outlined

in this guidance document is contained in this appendix, as are Internet addresses at which you can download no-cost software.

Appendix I - Examples of Planning, Implementation, and Assessment for RCRA Waste Sampling -- Two hypothetical examples of how to apply the planning, implementation, and assessment guidance provided in this guidance document are provided here.

Appendix J - Summaries of ASTM Standards -- This appendix provides summaries of ASTM standards related to waste sampling and referenced in this document.

2 SUMMARY OF RCRA REGULATORY DRIVERS FOR WASTE SAMPLING AND ANALYSIS

2.1 Background

Through RCRA, Congress provided EPA with the framework to develop regulatory programs for the management of solid and hazardous waste. The provisions of RCRA Subtitle C establish the criteria for identifying hazardous waste and managing it from its point of generation to ultimate disposal. EPA's regulations set out in 40 CFR Parts 260 to 279 are the primary source for the requirements of the hazardous waste program. These regulations were developed over a period of 25 years. While EPA's approach for developing individual regulations may have evolved over this period, the current RCRA statute and codified regulations remain the standard for determining compliance.

Many of the RCRA regulations either *require* the waste handler to conduct sampling and analysis, or they include provisions under which sampling and analysis can be performed at the discretion of the waste handler. If the regulations require sampling and analysis of a waste or environmental media, then any regulatory requirements for conducting the sampling and analysis and for evaluating the results must be followed. Regardless of whether there are regulatory requirements to conduct sampling, some waste handlers may wish to conduct a sampling program that allows them to quantify any uncertainties associated with their waste classification decisions. The information in this document can be used to aid in the planning and implementation of such a sampling program.

Some RCRA regulations *do not* specify sampling and analysis requirements and/or *do not* specify how the sample analysis results should be evaluated. In many cases, this is because EPA realized that the type, quantity, and quality of data needed should be specified on a site-specific basis, such as in the waste analysis plan of a permitted facility. In those situations, you can use the guidance in this document to help you plan and implement the sampling and analysis program, evaluate the sample analysis results against the regulatory standards, and quantify the level of uncertainty associated with the decisions.

This section identifies the major RCRA programs that specify some sort of sampling and testing to determine if a waste is a hazardous waste, to determine if a hazardous waste treatment standard is attained, or to meet other objectives such as site characterization. Table 1 provides a listing of these major RCRA programs that may require waste sampling and testing as part of their implementation. Appendix B provides a more detailed listing of the regulatory citations, the applicable RCRA standards, requirements for demonstrating attainment or compliance with the standards, and relevant USEPA guidance documents.

Prior to conducting a waste sampling and testing program to comply with RCRA, review the specific regulations in detail. Consult the latest 40 CFR, related *Federal Register* notices, and EPA's World Wide Web site (www.epa.gov) for new or revised regulations. In addition, because some states have requirements that differ from EPA regulations and guidance, we recommend that you consult with a representative from your State if your State is authorized to implement the regulation.

Table 1. Major RCRA Program Areas Involving Waste Sampling and Analysis [1]

40 CFR Citation	*Program Description*
Hazardous Waste Identification	
§ 261.3(a)(2)(v)	Used oil rebuttable presumption (also Part 279, Subparts B, E, F and G standards for the management of used oil)
§ 261.3(c)(2)(ii)(C)	Generic exclusion levels for K061, K062, and F006 nonwastewater HTMR residues
§ 261.21	Characteristic of Ignitability
§ 261.22	Characteristic of Corrosivity
§ 261.23	Characteristic of Reactivity
§ 261.24	Toxicity Characteristic
§ 261.38(c)(8)	Exclusion of Comparable Fuels from the Definition of Solid and Hazardous Waste
Part 261, Appendix I	Representative Sampling Methods
Mixed Hazardous Waste	Joint EPA-NRC sampling guidance. See November 20, 1997 *Federal Register* (62 *FR* 62079)
Land Disposal Restriction Program	
§ 268.6	Petitions to Allow Land Disposal of a Waste Prohibited Under Subpart C of Part 268 (No-Migration Petition). Sampling and testing criteria are specified at § 268.6(b)(1) and (2).
§ 268.40	Land Disposal Restriction (LDR) concentration-level standards
§ 268.44	Land Disposal Restriction Treatability Variance
§ 268.49(c)(1)	Alternative LDR Treatment Standards for Contaminated Soil
Other RCRA Programs and References	
§ 260.10	Definitions (for Representative Sample)
Part 260, Subpart C	Rulemaking Petitions
Part 262, Subpart A	Generator Standards - General (including § 262.11 Hazardous Waste Determination)
Part 262, Subpart C	Pre-Transport Requirements
Part 264, Subpart A	Treatment, Storage, and Disposal Facility Standards - General
Parts 264/265, Subpart B	Treatment, Storage, and Disposal Facility Standards - General Facility Standards
Parts 264/265, Subpart F	Releases from Solid Waste Management Units (ground-water monitoring)
Parts 264/265, Subpart G	Closure and Post-Closure
Parts 264, Subpart I	Use and Management of Containers
Parts 264/265 - Subpart J	Tank Systems

1. Expanded descriptions of the programs listed in Table 1 are given in Appendix B.

Table 1. Major RCRA Program Areas Involving Waste Sampling and Analysis (continued)

40 CFR Citation	*Program Description*
Other RCRA Programs and References (continued)	
Parts 264/265 - Subpart M	Land Treatment
Part 264/265 - Subpart O	Incinerators
Part 264, Subpart S	Corrective Action for Solid Waste Management Units (including § 264.552 Corrective Action Management Units)
Parts 264/265 - Subparts AA/BB/CC	Air Emission Standards
Part 266 - Subpart H	Hazardous Waste Burned in Boiler and Industrial Furnaces (BIFs) (including § 266.112 Regulation of Residues)
Part 270 - Subpart B	Permit Application, Hazardous Waste Permitting
Part 270 - Subpart C	Conditions Applicable to All Permits
Part 270 - Subpart F	Special Forms of Permits
Part 273	Standards for Universal Waste Management
Part 279	Standards for the Management of Used Oil

2.2 Sampling For Regulatory Compliance

Many RCRA programs involve sampling and analysis of waste or environmental media by the regulated community. Sampling and analysis often is employed to make a hazardous waste determination (see Section 2.2.1), to determine if a waste is subject to treatment or, if so, has been adequately treated under the Land Disposal Restrictions program (see Section 2.2.2), or in responding to other RCRA programs that include routine monitoring, unit closure, or cleanup (see Section 2.2.3).

2.2.1 Making a Hazardous Waste Determination

Under RCRA, a hazardous waste is defined as a solid waste, or a combination of solid wastes which, because of its quantity, concentration, or physical, chemical, or infectious characteristics, may cause, or significantly contribute to an increase in mortality or an increase in serious irreversible or incapacitating reversible illness, or pose a substantial present or potential hazard to human health or the environment when improperly treated, stored, transported, disposed, or otherwise managed. The regulatory definition of a hazardous waste is found in 40 CFR § 261.3.

Solid wastes are defined by regulation as hazardous wastes in two ways. First, solid wastes are hazardous wastes if EPA lists them as hazardous wastes. The lists of hazardous wastes are found in 40 CFR Part 261, Subpart D. Second, EPA identifies the characteristics of a hazardous waste based on criteria in 40 CFR § 261.10. Accordingly, solid wastes are hazardous if they exhibit any of the following four characteristics of a hazardous waste: ignitability, corrosivity, reactivity, or toxicity (based on the results of the Toxicity Characteristic Leaching Procedure, or TCLP). Descriptions of the hazardous waste characteristics are found in 40 CFR Part 261, Subpart C.

Generators must conduct a hazardous waste determination according to the hierarchy specified in 40 CFR § 262.11. Persons who generate a solid waste first must determine if the solid waste is excluded from the definition of hazardous waste under the provisions of 40 CFR § 261.4. Once the generator determines that a solid waste is not excluded, then he/she must determine if the waste meets one or more of the hazardous waste listing descriptions and determine whether the waste is mixed with a hazardous waste, is derived from a listed hazardous waste, or contains a hazardous waste.

For purposes of compliance with 40 CFR Part 268, or if the solid waste is not a listed hazardous waste, the generator must determine if the waste exhibits a characteristic of a hazardous waste. This evaluation involves testing the waste *or* using knowledge of the process or materials used to produce the waste.

When a waste handler conducts testing to determine if the waste exhibits any of the four characteristics of a hazardous waste, he or she must obtain a representative sample (within the meaning of a representative sample given at § 260.10) using the applicable sampling method specified in Appendix I of Part 261 or alternative method (per § 261.20(c))[1] and test the waste for the hazardous waste characteristics of interest at § 261.21 through 261.24.

For the purposes of subpart 261, the identification of hazardous waste, the regulations state that a sample obtained using any of the applicable sampling methods specified in Appendix I of Part 261 to be a representative sample within the meaning of the Part 260 definition of representative sample. Since these sampling methods are not officially required, anyone desiring to use a different sampling method may do so without demonstrating the equivalency of that method under the procedures set forth in § 260.21. The user of an alternate sampling method must use a method that yields samples that "meet the definition of representative sample found in Part 260" (45 FR 33084 and 33108, May 18, 1990). Such methods should enable one to obtain samples that are equally representative as those specified in Appendix I of Part 261. The planning process and much of the information described in this guidance document may be helpful to someone regulated under Part 261 wishing to use an alternate sampling method. The guidance should be help full as well for purposes other than Part 261.

Certain states also may have requirements for identifying hazardous wastes in addition to those requirements specified by Federal regulations. States authorized to implement the RCRA or HSWA programs under Section 3006 of RCRA may promulgate regulations that are more stringent or broader in scope than Federal regulations.

2.2.2 Land Disposal Restrictions (LDR) Program

The LDR program regulations found at 40 CFR Part 268 require that a hazardous waste generator determine if the waste has to be treated before it can be land disposed. This is done by determining if the hazardous waste meets the applicable treatment standards at § 268.40, § 268.45, or §268.49. EPA expresses treatment standards either as required treatment technologies that must be applied to the waste or as contaminant concentration levels that must

[1] Since the 40 CFR Part 261 Appendix I sampling methods are not formally adopted by the EPA Administrator, a person who desires to employ an alternative sampling method is not required to demonstrate the equivalency of his or her method under the procedures set forth in §§ 260.20 and 260.21 (see comment at § 261.20(c)).

be met. (Alternative LDR treatments standards have been promulgated for contaminated soil, debris, and lab packs.) Determining the need for waste treatment can be made by either of two ways: testing the waste or using knowledge of the waste (see § 268.7(a)).

If a hazardous waste generator is managing and treating prohibited waste or contaminated soil in tanks, containers, or containment buildings to meet the applicable treatment standard, then the generator must develop and follow a written waste analysis plan (WAP) in accordance with § 268.7(a)(5).

A hazardous waste treater must test their waste according to the frequency specified in their WAP as required by 40 CFR 264.13 (for permitted facilities) or 40 CFR 265.13 (for interim status facilities). See § 268.7(b).

If testing is performed, *no portion of the waste may exceed the applicable treatment standard*, otherwise, there is evidence that the standard is not met (see 63 FR 28567, March 26, 1998). Statistical variability is “built in” to the standards (USEPA 1991c). Wastes that do not meet treatment standards can not be land disposed unless EPA has granted a variance, extension, or exclusion (or the waste is managed in a "no-migration unit"). In addition to the disposal prohibition, there are prohibitions and limits in the LDR program regarding the dilution and storage of wastes. The program also requires tracking and recordkeeping to ensure proper management and safe land disposal of hazardous wastes.

General guidance on the LDR program can be found in *Land Disposal Restrictions: Summary of Requirements* (USEPA 2001d). Detailed guidance on preparing a waste analysis plan (WAP) under the LDR program can be found in *Waste Analysis at Facilities That Generate, Treat, Store, and Dispose of Hazardous Wastes - A Guidance Manual* (USEPA 1994a). Detailed guidance on measuring compliance with the alternative LDR treatment standards for contaminated soil can be found in *Guidance on Demonstrating Compliance With the Land Disposal Restrictions (LDR) Alternative Soil Treatment Standards* (USEPA 2002a).

2.2.3 Other RCRA Regulations and Programs That May Require Sampling and Testing

In addition to the RCRA hazardous waste identification regulations and the LDR regulations, EPA has promulgated other regulations and initiated other programs that may involve sampling and testing of solid waste and environmental media (such as ground water or soil). Program-specific EPA guidance should be consulted prior to implementing a sampling or monitoring program to respond to the requirements of these regulations or programs. For example, EPA has issued separate program-specific guidance on sampling to support preparation of a delisting petition, ground-water and unsaturated zone monitoring at regulated units, unit closure, corrective action for solid waste management units, and other programs. See also Appendix B of this document.

2.2.4 Enforcement Sampling and Analysis

The sampling and analysis conducted by a waste handler during the normal course of operating a waste management operation might be quite different than the sampling and analysis conducted by an enforcement agency. The primary reason is that the data quality objectives (DQOs) of the enforcement agency often may be legitimately different from those of a waste handler. Consider an example to illustrate this potential difference in approach: Many of

RCRA's standards were developed as concentrations that should not be exceeded (or equaled) or as characteristics that should not be exhibited for the waste or environmental media to comply with the standard. In the case of such a standard, the waste handler and enforcement officials might have very different objectives. An enforcement official, when conducting a compliance sampling inspection to evaluate a waste handler's compliance with a "do not exceed" standard, take only one sample. Such a sample may be purposively selected based on professional judgment. This is because all the enforcement official needs to observe – for example to determine that a waste is hazardous – is a single exceedance of the standard.

A waste handler, however, in responding to the same regulatory standard may want to ensure, with a specified level of confidence, that his or her waste concentrations are low enough so that it would be unlikely, for example, that an additional sample drawn from the waste would exceed the regulatory standard. In designing such an evaluation the waste handler could decide to take a sufficient number of samples in a manner that would allow evaluation of the results statistically to show, with the desired level of confidence, that there is a low probability that another randomly selected sample would exceed the standard.

An important component of the enforcement official's DQO is to "prove the positive." In other words, the enforcement official is trying to demonstrate whether the concentration of a specific constituent in some portion of the waste exceeds the "do not exceed" regulatory level. The "prove the positive" objective combined with the "do not exceed" standard only requires a single observation above the regulatory level in order to draw a valid conclusion that at least some of the waste exceeds the level of concern.

The Agency has made it clear that in "proving the positive," the enforcement agency's DQOs may not require low detection limits, high analyte recoveries, or high degrees of precision:

> *"If a sample possesses the property of interest, or contains the constituent at a high enough level relative to the regulatory threshold, then the population from which the sample was drawn must also possess the property of interest or contain that constituent. Depending on the degree to which the property of interest is exceeded, testing of samples which represent all aspects of the waste or other material may not be necessary to prove that the waste is subject to regulation"* (see 55 FR 4440, "Hazardous Waste Management System: Testing and Monitoring Activities," February 8, 1990).

A waste handler may have a different objective when characterizing his or her waste. Instead, the waste handler may wish to "prove the negative." While proving the negative in absolute terms is not realistic, the waste handler may try to demonstrate with a desired level of confidence that the vast majority of his or her waste is well below the standard such that another sample or samples taken from the waste would not likely exceed the regulatory standard. The Agency also has spoken to the need for sound sampling designs and proper quality control when one is trying to "prove the negative:"

> *"The sampling strategy for these situations (proving the negative) should be thorough enough to insure that one does not conclude a waste is nonhazardous when, in fact, it is hazardous. For example, one needs to take enough samples so that one does not miss areas of high concentration in an otherwise clean material. Samples must be handled so that properties do not change and*

> *contaminants are not lost. The analytical methods must be quantitative, and regulatory detection limits must be met and documented"* (see 55 FR 4440, "Hazardous Waste Management System: Testing and Monitoring Activities," February 8, 1990).

"Proving the negative" can be a more demanding objective for the waste handler in terms of the sampling strategy and resources than that faced by the enforcement official. To address this objective the waste handler could use the advice in this or similar guidance documents. In doing so, the waste handler should establish objectives using a systematic planning process, design a sampling and analysis plan based on the objectives, collect and analyze the appropriate number of samples, and use the information from the sample analysis results for decision-making.

The distinction between a sampling strategy designed to "prove the negative" versus one designed to "prove the positive" also has been supported in a recent judicial ruling. In *United States v. Allen Elias* (9th Cir. 2001) the Government used a limited number of samples to prove that hazardous waste was improperly managed and disposed. The court affirmed that additional sampling by the Government was not necessary to "prove the positive."

3 FUNDAMENTAL STATISTICAL CONCEPTS

Throughout the life cycle of a waste-testing program, the tools of statistics often are employed -- in planning, implementation, and assessment. For example, in the planning phase, you may state certain project objectives quantitatively and use statistical terminology. Designing and implementing a sampling plan requires an understanding of error and uncertainty. Statistical techniques can be used to describe and evaluate the data and to support decisions regarding the regulatory status of a waste or contaminated media, attainment of treatment or cleanup goals, or whether there has been a release to the environment. Because statistical concepts may be used throughout the sampling and analysis program, an understanding of basic statistical concepts and terminology is important.

While statistical methods can be valuable in designing and implementing a scientifically sound waste-sampling program, their use should not be a substitute for knowledge of the waste or as a substitute for common sense. Not every problem can, or necessarily must, be evaluated using probabilistic techniques. Qualitative expressions of decision confidence through the exercise of professional judgment (such as a "weight of evidence" approach) may well be sufficient, and in some cases may be the only option available (Crumbling 2001).

Do the RCRA regulations require statistical sampling?

Some RCRA regulations *require* the use of statistical tests (e.g., to determine if there has been a release to ground water from a waste management unit under 40 CFR Subpart F), whereas, other RCRA regulations *do not* require the use of statistical tests (such as those for determining if a solid waste is or is not a hazardous waste or determining compliance with LDR treatment standards). Even where there is no regulatory obligation to conduct sampling or apply statistical tests to evaluate sampling results, statistical methods can be useful in interpreting data and managing uncertainty associated with waste classification decisions.

If the objective of the sampling program is to make a hazardous waste determination, the regulations allow that a single representative sample is sufficient to classify a waste as hazardous. If a representative sample is found to have the properties set forth for the corrosivity, ignitability, reactivity, or toxicity characteristics, then the waste is hazardous. The regulations do not address directly what is a sufficient number of samples to classify a solid waste as *nonhazardous*. However, for a petition to reclassify (delist) a listed hazardous waste, which includes a determination that the listed hazardous waste is not a characteristic hazardous waste (a "nonhazardous" classification), the regulations provide that at least four representative samples sufficient to represent the variability or uniformity of the waste must be tested (40 CFR 260.22). This approach is not necessarily based on any statistical method but reflects concepts of proving the negative and proving the positive (see also Section 2.2.4).

Even if you have no formal training in statistics, you probably are familiar with basic statistical concepts and how samples are used to make inferences about the population from which the samples were drawn. For example, the news media frequently cite the results of surveys that make generalized conclusions about public opinion based on interviews with a relatively small proportion of the population. These results, however, are only *estimates* because no matter how carefully a survey is done, if repeated over and over in an identical manner, the answer will be a little different each time. There always will be some random sampling variation because it is not possible to survey every member of a population. There also will be measurement and estimation errors because of mistakes made in how data are obtained and interpreted. Responsible pollsters report this as their "margin of error" along with the findings of the survey

(Edmondson 1996).

Similar to surveys of human populations, waste characterization studies can be designed in such a way that a population can be identified, samples can be collected, and the uncertainty in the results can be reported.

The following sections provide a brief overview of the statistical concepts used in this guidance. Four general topics are described:

- Populations, samples, and distributions (Section 3.1)
- Measures of central tendency, variability, and relative standing (Section 3.2)
- Precision and bias (Section 3.3)
- Using sample analysis results to classify a waste or determine its status under RCRA (Section 3.4).

Guidance on selecting and using statistical methods for evaluating data is given in Section 8.2 and Appendix F of this document. Statistical tables are given in Appendix G. Additional statistical guidance can be found in *Guidance for Data Quality Assessment, EPA QA/G-9* (USEPA 2000d) and other references cited.

3.1 Populations, Samples, and Distributions

A "population" consists of all the waste or media whose characteristics are to be studied and estimated. A set of observations, known as a statistical sample, is a portion of the population that is studied in order to learn about the whole population. Sampling is necessary when a study of the entire population would be too expensive or physically impossible.

Inferences about the population are made from samples selected from the population. For example, the sample mean (or average) is a consistent estimator of the population mean. In general, estimates made from samples tend to more closely approximate the true population parameter as the number of samples increases. The precision of these inferences depends on the theoretical sampling distribution of the statistic that would occur if the sampling process were repeated over and over using the same sampling design and number of samples.

3.1.1 Populations and Decision Units

A "population" is the *entire* selection of interest for study. Populations can have *spatial* boundaries, which define the physical area to be studied, and *temporal* boundaries, which describe the time interval the study will represent. The definition of the population can be subjective, defined by regulation or permit condition, or based on risks to human health and the environment. In all cases, however, the population needs to be finite and have well-defined, unambiguous physical and/or temporal boundaries. The physical boundary defines the size, shape, orientation, and location of the waste or media about which a decision will be made.

For a large population of waste or media, you may wish to subdivide the population into smaller units about which decisions can be made, rather than attempt to characterize the entire

population. These units are called "decision units," and they may represent a single type of waste at the point of waste generation, a waste from a single batch operation, waste generated over a specified time, or a volume of waste or contaminated media (such as soil) subject to characterization, removal, and/or treatment. The concept of a decision unit is similar to an "exposure unit" (Neptune, et al. 1990, Blacker and Goodman 1994a and 1994b, Myers 1997), or "exposure area" (USEPA 1992a and 1996a) in EPA's Superfund program in which risk-based decisions consider the mass or area of the waste or media. A decision unit also is analogous to a "remediation unit" as described in EPA's *Data Quality Objective Process for Superfund* (USEPA 1993a).

When using samples to determine whether a solid waste is a hazardous waste, that determination must be made at the **point of generation** (i.e., when the waste becomes a solid waste).

Hypothetical examples of populations or decision units that might be encountered in the context of RCRA waste characterization follow:

- Filter cake being placed in a 25-cubic-yard roll-off bin at the point of waste generation
- Waste water contained in a 55-gallon drum
- Liquid waste flowing from the point of generation during a specified time interval
- A block of soil (e.g., 10-feet-by-10-feet square, 6-inches deep) within a solid waste management unit (SWMU).

In some situations, it will be appropriate to define two separate populations for comparison to each other. For example, in monitoring a land-based waste management unit to determine if there has been a release to the subsurface at statistically significant levels above background, it is necessary to establish two populations: (1) a background population and (2) an exposed (or downgradient) population in the soil, pore-water, or ground-water system.

In situations in which the boundaries of the waste or contamination are not obvious or cannot be defined in advance (such as the case of contaminated soil *in situ*, as opposed to excavated soil in a pile), the investigator is interested in the *location* of the contamination as well as the concentration information. Such a sampling objective is best addressed by spatial analysis, for example, by using geostatistical methods (See also Section 3.4.4).

3.1.2 Samples and Measurements

Samples are portions of the population. Using information from a set of samples (such as measurements of chemical concentrations) and the tools of inductive statistics, inferences can be made about the population. The validity of the inferences depends on how closely the samples represent the physical and chemical properties of the population of interest.

In this document, we use the word "sample" in several different ways. To avoid confusion, definitions of terms follow:

Sample: A portion of material that is taken from a larger quantity for the purpose of estimating properties or composition of the larger quantity (from ASTM D 6233-98).

Statistical sample: A set of samples or measurements selected by probabilistic means (i.e., by using some form of randomness).

We sometimes refer to a "set of samples" to indicate more than one individual sample that may or may not have been obtained by probabilistic means.

Outside the fields of waste management and environmental sciences, the concept of a sample or "sampling unit" is fairly straightforward. For example, a pollster measures the opinions of individual human beings, or the QC engineer measures the diameter of individual ball bearings. It is easy to see that the measurement and the sampling unit correspond; however, in sampling waste or environmental media, *what is the appropriate "portion" that should be in a sampling unit*? The answer to this question requires consideration of the heterogeneities of the sample media and the dimension of the sampling problem (in other words, are you sampling over time or sampling over space?). The information can be used to define the appropriate *size, shape, and orientation* of the sample. The size, shape, and orientation of a sample are known as the **sample support**, and the sample support will affect the measurement value obtained from the sample.

As shown in Figure 2, after a sample of a certain *size, shape, and orientation* is obtained in the field (as the primary sample), it is handled, transported, and prepared for analysis. At each stage, changes can occur in the sample (such as the gain or loss of constituents, changes in the particle size distribution, etc.). These changes accumulate as errors throughout the sampling process such that measurements made on relatively small analytical samples (often less than 1 gram) may no longer "represent" the population of interest. Because sampling and analysis results may be relied upon to make decisions about a waste or media, it is important to understand the sources of the errors introduced at each stage of sampling and take steps to minimize or control those errors. In doing so, samples will be sufficiently "representative" of the population from which they are obtained.

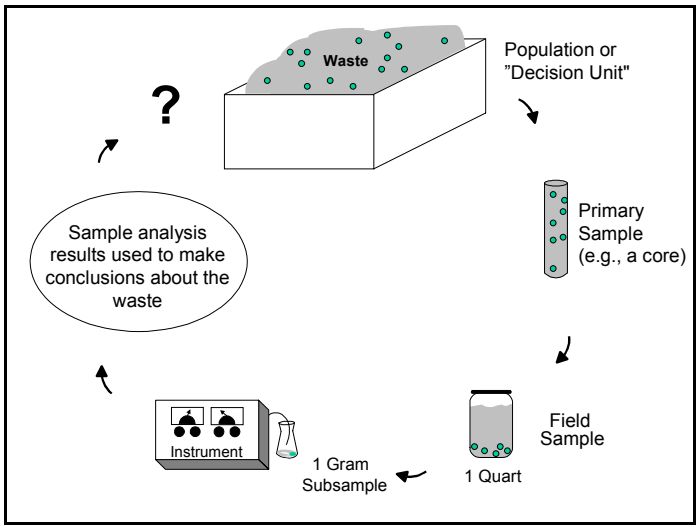

Figure 2. Very small analytical samples are used to make decisions about much larger volumes (modified after Myers 1997).

The RCRA solid waste regulations at 40 CFR §260.10 define a **representative sample** as:

> "a sample of a universe or whole (e.g., waste pile, lagoon, ground water) which can be expected to exhibit the average properties of the universe or whole."

RCRA implementors, at a minimum, must use this definition when a representative sample is called for by the regulations. Various other definitions of a representative sample have been developed by other organizations. For example, ASTM in their consensus standard D 6044-96 defines a representative sample as “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 was collected" (ASTM D 6044). A detailed discussion of representativeness also is given in *Guidance on Data Quality Indicators* (USEPA 2001e).

3.1.3 Distributions

Because the concentration of constituents of concern will not be the same for every individual sample, there must be a *distribution* of concentrations among the population. Understanding the distributional characteristics of a data set is an important first step in data analysis.

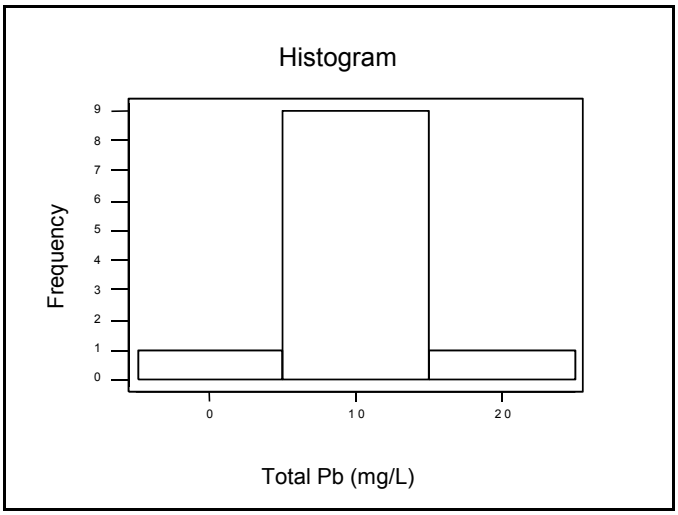

Figure 3. Histogram representing the distribution of total lead (Pb) in 11 samples of No. 2 fuel oil (USEPA 1998b).

If we have a sufficient number of samples selected from a population, a picture of the distribution of the sample data can be represented in the form of a **histogram**. A histogram, which offers a simple graphical representation of the shape of the distribution of data, can be constructed by dividing the data range into units or “bins” (usually of equal width), counting the number of points within each unit, and displaying the data as the height or area within a bar graph. Figure 3 is an example of a histogram made using analysis results for total lead in 11 samples of No. 2 fuel oil (data set from USEPA 1998b). Guidance on constructing histograms can be found in EPA’s *Guidance for Data Quality Assessment, EPA QA/G*-9 (USEPA 2000d).

With a sufficiently large number of samples, the bars of the histogram could be “blended together” to form a curve known as a probability density function (PDF). Figure 4 shows two probability density functions you might encounter: Figure 4(a) is a **normal distribution** with its familiar symmetrical mound-shape. Figure 4(b) is a **lognormal distribution** in which the natural log-transformed values exhibit a normal distribution. A lognormal distribution indicates that a relatively small proportion of the population includes some relatively large values.

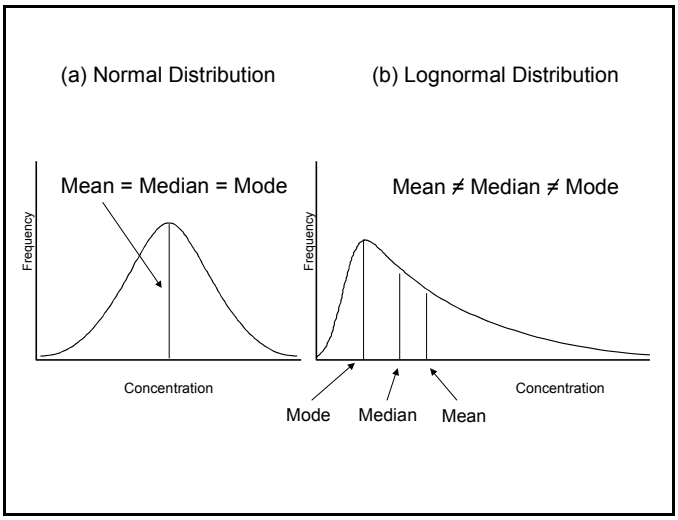

Figure 4. Examples of two distributions: (a) normal distribution and (b) lognormal distribution

Many of the tools used in statistics are based on the assumption that the data are normally distributed, can be transformed to a normal scale, or can be treated as if they are approximately normal. The assumption of a normal distribution often can be made without significantly increasing the risk of making a "wrong" decision. Of course, the normal and lognormal distributions are *assumed* models that only approximate the underlying population distribution.

Another distribution of interest is known as the **binomial distribution**. The binomial distribution can be used when the sample analysis results are interpreted as either "fail" or "pass" (e.g., a sample analysis result either exceeds a regulatory standard or does not exceed the standard).

In some cases, you may not be able to "fit" the data to any particular distributional model. In these situations, we recommend you consider using a "distribution-free" or "nonparametric" statistical method (see Section 8.2).

A simple but extremely useful graphical test for normality is to graph the data as a **probability plot**. In a probability plot, the vertical axis has a probability scale and the horizontal axis has a data scale. In general, if the data plot as a straight line, there is a qualitative indication of normality. If the natural logarithms of the data plot as a straight line, there is an indication of lognormality.

Figure 5 provides an example of a normal probability plot created from the same data used to generate the histogram in Figure 3. Guidance on constructing probability plots can be found in EPA's *Guidance for Data Quality Assessment*, *EPA QA/G-9* (USEPA 2000d).

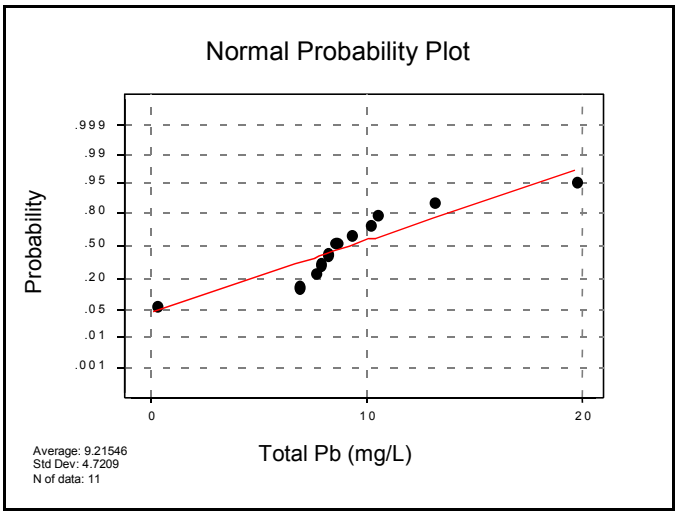

Figure 5. Normal probability plot

Section 8 (Assessment: Analyzing and Interpreting Data) provides guidance on checking the distribution of data sets and provides strategies for handling sample data exhibiting a non-normal distribution.

3.2 Measures of Central Tendency, Variability, and Relative Standing

In addition to graphical techniques for summarizing and describing data sets, numerical methods can be used. Numerical methods can be used to describe the central tendency of the set of measurements, the variability or spread of the data, and the relative standing or relative location of a measurement within a data set.

3.2.1 Measures of Central Tendency

The average or **mean** often is used as a measure of central tendency. The mean of a set of quantitative data is equal to the sum of the measurements divided by the number of measurements contained in the data set. Other measures of central tendency include the

median (the midpoint of an ordered data set in which half the values are below the median and half are above) and the **mode** (the value that occurs most often in the distribution). For distributions that are not symmetrical, the median and the mean do not coincide. The mean for a lognormal distribution, for instance, will exceed its median (see Figure 4(b)).

The true **population mean**, μ ("mu"), is the average of the true measurements (e.g., of the constituent concentration) made over all possible samples. The population mean is never known because we cannot measure all the members of a population (or all possible samples). We can, however, *estimate* the population mean by taking random samples from the population. The average of measurements taken on random samples is called the **sample mean**. The sample mean is denoted by the symbol $\bar{x}$ ("x-bar") and calculated by summing the value obtained from each random sample (x_i) and dividing by the number of samples (n):

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \textbf{Equation 1}$$

Box 1 provides an example calculation of the sample mean.

Box 1. Example Calculation of the Sample Mean

Using Equation 1 and the following four data points in parts per million (ppm): 86, 90, 98, and 104, the following is an example of computing the sample mean.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{86+90+98+104}{4} = 95 ppm$$

Therefore, the sample mean is 95 ppm.

3.2.2 Measures of Variability

Random variation in the population is described by "dispersion" parameters -- the **population variance** (σ^2) and the **population standard deviation** (σ). Because we cannot measure all possible samples that comprise the population, the values for σ^2 and σ are unknown. The variance, however, can be *estimated* from a statistical sample of the population by the **sample variance**:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad \textbf{Equation 2}$$

The variance calculated from the samples is known as the **sample variance** (s^2) and it includes random variation in the population as well as random variation that can be introduced by sample collection and handling, sample transport, and sample preparation and analysis. The sample variance is an estimate of the variance that one would obtain if the entire set of all possible samples in the population were measured using the same measurement process as is

being employed for the n samples. If there were no sample handling or measurement error, this sample variance (s^2) would estimate the population variance (σ^2).

The **population standard deviation** (σ) is estimated by s, the **sample standard deviation**:

$$s = \sqrt{s^2} \qquad \textbf{Equation 3}$$

Box 2 provides an example calculation of the sample variance and sample standard deviation.

Box 2. Example Calculations of Sample Variance and Standard Deviation

Using Equation 2 and the data points in Box 1, the following is an example calculation of the sample variance:

$$s^2 = \frac{\left[(86-94.5)^2 + (90-94.5)^2 + (98-94.5)^2 + (104-94.5)^2\right]}{4-1} = \frac{195}{3} = 65$$

Using Equation 3, the sample standard deviation is then calculated as follows:

$$s = \sqrt{s^2} = 8.1$$

The standard deviation is used to measure the variability in a data set. For a normal distribution, we know the following (see Figure 6):

- Approximately 68 percent of measurements will fall within $\pm$ 1 standard deviation of the mean
- Approximately 95 percent of the measurements will fall within $\pm$ 2 standard deviations of the mean
- Almost all (99.74 percent) of the measurements will fall within $\pm$ 3 standard deviations of the mean.

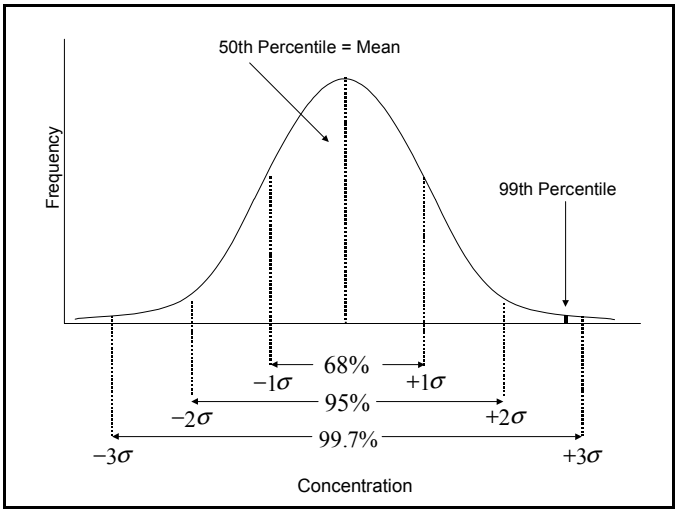

Figure 6. Percentage of values falling within 1, 2, and 3 standard deviations of the mean of a normal distribution. The figure also shows the relationship between the mean, the 50th percentile, and the 99th percentile in a normal distribution.

Estimates of the standard deviation, combined with the assumption of a normal distribution, allow us to make quantitative statements about the spread of the data. The larger the spread in the data, the less certainty we have in estimates or decisions made from the data. As discussed in the following section, a small spread in the data offers

more certainty in estimates and decisions made from the data.

Because $\bar{x}$ is an estimate of a population parameter based on a statistical sample, we expect its value to be different each time a new set of samples is drawn from the population. The means calculated from repeated statistical samples also form a distribution. The estimate of the standard deviation of the sampling distribution of means is called the **standard error**.

The standard error of the mean ($s_{\bar{x}}$) is estimated by:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \qquad \textbf{Equation 4}$$

The standard error is used in equations to calculate the appropriate number of samples to estimate the mean with specified confidence (see Section 5.4), and it is used in statistical tests to make inferences about $\bar{x}$ (see Appendix F).

3.2.3 Measures of Relative Standing

In addition to measures of central tendency and variability to describe data, we also may be interested in describing the relative standing or location of a particular measurement within a data set. One such measure of interest is the **percentile** ranking. A population percentile represents the percentage of elements of a population having values less than a specified value. Mathematically, for a set of n measurements the $p\text{th}$ percentile (or quantile) is a number such that $p\%$ of the measurements fall below the $p\text{th}$ percentile, and $(100-p)\%$ fall above it. For example, if a measurement is located at the 99th percentile in a data set, it means that 99 percent of measurements are less than that measurement, and 1 percent are above. In other words, almost the *entire* distribution lies below the value representing the 99th percentile. Figure 6 depicts the relationship between the mean, the 50th percentile, and the 99th percentile in a normal distribution.

Just like the mean and the median, a percentile is a population parameter that must be estimated from the sample data. As indicated in Figure 6, for a normal distribution a "point estimate" of a percentile ($\hat{x}_p$) can be obtained using the sample mean ($\bar{x}$) and the sample standard deviation (s) by:

$$\hat{x}_p = \bar{x} + z_p s \qquad \textbf{Equation 5}$$

where z_p is the $p\text{th}$ quantile of the standard normal distribution. (Values of z_p that correspond to values of p can be obtained from the last row of Table G-1 in Appendix G). A probability plot (see Figure 5) offers another method of estimating normal percentiles. See EPA's *Guidance for Data Quality Assessment, EPA QA/G-9* (USEPA 2000d) for guidance on constructing probability plots and estimating percentiles.

3.3 Precision and Bias

The representativeness of a statistical sample (that is, a set of samples) can be described in terms of **precision** and **bias**. Precision is a measurement of the *closeness of agreement* between repeated measurements. Bias is the systematic or consistent over- or underestimation of the true value (Myers 1997, USEPA 2000d).

The analogy of a target often is used to illustrate the concepts of precision and bias. In Figure 7, the center of each target represents the true (but unknown) average concentration in a batch of waste. The "shots" in targets (a) through (d) represent measurement results from samples taken to estimate the true concentration. The figure also can be used to illustrate precision and bias associated with measurement processes within a laboratory in which the same sample is analyzed multiple times (for example, four times).

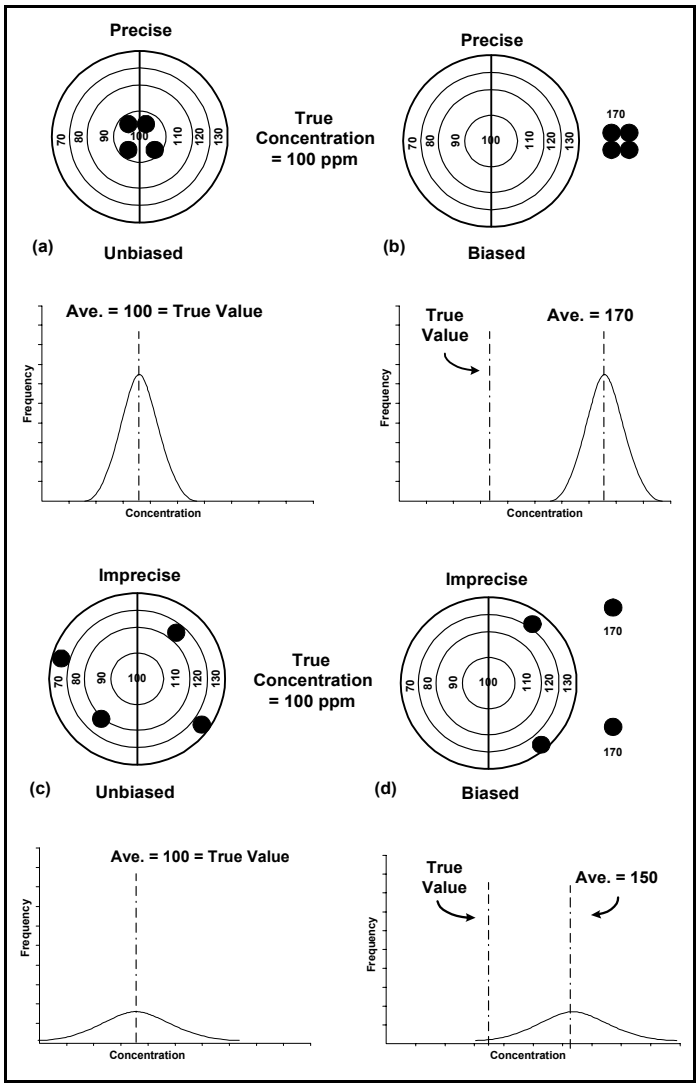

Figure 7. Shots at a target illustrate precision and bias (modified after Jessen 1978).

Figure 7(a) indicates high precision and low bias in the sampling and analysis results. Generally, high precision and minimal bias are required when one or more chemical constituents in a solid waste are present at concentrations close to the applicable regulatory threshold or action level. Note that each of the measurements in Figure 7(a) is in close agreement with the true value. These measurements can be described as having high **accuracy**.

If the sampling and measurement process is very precise but suffers from bias (such as use of an incorrect sampling procedure or contamination of an analytical instrument), the situation could be as pictured in Figure 7(b) in which the repeated measurements are close to one another but not close to the true value. In fact, the data express a significant 70 percent bias that might go undetected if the true value is not known.

The opposite situation is depicted in Figure 7(c), where the data show low precision (that is, high dispersion around the mean) but are unbiased because the samples lack any systematic error and the average of the measurements reflects the true average concentration. Precision in sampling can be improved by increasing the number of samples, increasing the volume

(mass) of each sample, or by employing a composite sampling strategies. Note, however, that relatively imprecise results can be tolerated if the contaminants of concern occur at levels either far below or far above their applicable thresholds.

Figure 7(d) depicts the situation where the sampling and analytical process suffers from both imprecision and bias. In both Figures 7(b) and (d), the bias will result in an incorrect estimate of the true concentration, even if innumerable samples are collected and analyzed to control the impact of imprecision (i.e., bias will not "cancel out" with increasing numbers of samples).

There are several types and causes of bias, including sampling bias, analytical bias, and statistical bias:

> **Sampling Bias:** There are three potential sources of sampling bias: (1) Bias can be introduced in the field and the laboratory through the improper selection and use of devices for sampling and subsampling. Bias related to sampling tools can be minimized by ensuring all of the material of interest for the study is accessible by the sampling tool. (2) Bias can be introduced through improper design of the sampling plan. Improper sampling design can cause parts of the population of interest to be over- or under-sampled, thereby causing the estimated values to be systematically shifted away from the true values. Bias related to sampling design can be minimized by ensuring the sampling protocol is impartial so there is an equal chance for each part of the waste to be included in the sample over both the spatial and temporal boundaries defined for the study. (3) Bias can be introduced in sampling due to the loss or addition of contaminants during sampling and sample handling. This bias can be controlled using sampling devices made of materials that do not sorb or leach constituents of concern, and by use of careful decontamination and sample handling procedures. For example, agitation or homogenization of samples can cause a loss of volatile constituents, thereby indicating a concentration of volatiles lower than the true value. Proper decontamination of sampling equipment between sample locations or the use of disposable devices, and the use of appropriate sample containers and preservatives also can control bias in field sampling.
>
> **Analytical Bias:** Analytical (or measurement) bias is a systematic error caused by instrument contamination, calibration drift, or by numerous other causes, such as extraction inefficiency by the solvent, matrix effect, and losses during shipping and handling.
>
> **Statistical Bias:** After the sample data have been obtained, statistics are used to estimate population parameters using the sample data. Statistical bias can occur in two situations: (1) when the assumptions made about the sampling distribution are not consistent with the underlying population distribution, or (2) when the statistical estimator itself is biased.

Returning to Figure 7, note that each target has an associated frequency distribution curve. Frequency curves are made by plotting a concentration value versus the frequency of occurrence of that concentration. The curves show that as precision decreases (i.e., the variance σ^2 increases), the curve flattens out and an increasing number of measurements are found further away from the average (figures c and d). More precise measurements result in steeper curves (figures a and b) with the majority of measurements relatively closer to the

average value in normally distributed data. The greater the bias (figures b and d) the further the average of the measurements is shifted away from the true value. The smaller the bias (figures a and c) the closer the average of the samples is to the true average.

Representative samples are obtained by controlling (at acceptable levels) random variability (σ^2) and systematic error (or bias) in sampling and analysis. Quality control procedures and samples are used to estimate the precision and bias of sampling and analytical results.

3.4 Using Sample Analysis Results to Classify a Waste or to Determine Its Status Under RCRA

If samples are used to classify a waste or determine its regulatory status, then the sampling approach (including the number and type of samples) must meet the requirements specified by the regulations. Regardless of whether or not the regulations specify sampling requirements or the use of a statistical test, the Agency encourages waste handlers to use a systematic planning process such as the DQO Process to set objectives for the type, quantity, and quality of data needed to ensure with some known level of assurance that the regulatory standards are achieved.

After consideration of the objectives identified in the planning process, careful implementation of the sampling plan, and review of the analytical results, you can use the sample analysis results to classify a waste or make other decisions regarding the status of the waste under RCRA. The approach you select to obtain and evaluate the results will be highly dependent on the regulatory requirements (see Section 2 and Appendix B) and the data quality objectives (see Section 4 and Section 5).

The following sections provide a conceptual overview of how you can use sample analysis results to classify a waste or determine its status under RCRA. Guidance is provided on the following topics:

- Using an *average* to measure compliance with a fixed standard (Section 3.4.1)
- Using the *maximum* sample analysis result or an upper *percentile* to measure compliance with a fixed standard (Section 3.4.2)

There are other approaches you might use to evaluate sample analysis results, including tests that compare two populations, such as “downgradient” to “background” (see Section 3.4.3), and analysis of spatial patterns of contamination (see Section 3.4.4).

Detailed statistical guidance, including the necessary statistical equations, is provided in Section 8.2 and Appendix F.

3.4.1 Using an Average To Determine Whether a Waste or Media Meets the Applicable Standard

The arithmetic average (or mean) is a common parameter used to determine whether the concentration of a constituent in a waste or media is below a fixed standard. The mean often is used in cases in which a long-term (chronic) exposure scenario is assumed (USEPA 1992c) or where some average condition is of interest.

Because of the uncertainty associated with estimating the true mean concentration, a **confidence interval on the mean** is used to define the upper and lower limits that bracket the true mean with a known level of confidence. If the **upper confidence limit** (UCL) on the mean is less than the fixed standard, then we can conclude the true average is below the standard with a known amount of confidence. As an alternative to using a statistical interval to draw conclusions from the data, you could use hypothesis testing as described in EPA's *Guidance for the Data Quality Objectives Process, EPA QA/G-4* (USEPA 2000b) *and Guidance for Data Quality Assessment, EPA QA/G-9* (USEPA 2000d).

Confidence intervals are calculated using the sample analysis results. Figure 8 shows what is expected to happen when ten different sets of samples are drawn from the same waste and a confidence interval for the mean is calculated for each set of samples. The true (but unknown) mean (μ) – shown as a vertical line – does not change, but the positions of the *sample means* ($\bar{x}$) and confidence intervals (shown as the horizontal lines) do change. For most of the sampling events, the confidence interval contains the true mean, but sometimes it does not. In this particular example, we expect 8 out of 10 intervals to contain the true mean, so we call this an "80-percent confidence interval on the mean." In practice, you only have one set of data from one sampling event, not ten. Note that an equal degree of uncertainty is associated with the parameter of interest being located outside each of the two interval endpoints. Consequently, the confidence interval employed in this example is, for all practical purposes, a 90-percent interval. We will refer to this as a "one-sided 90-percent confidence limit on the mean." Of course, other levels of confidence could be used, such as a 95-percent confidence limit.

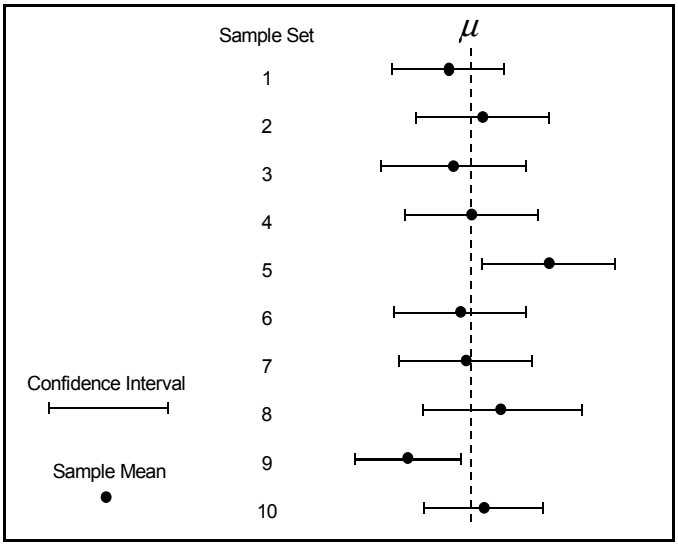

Figure 8. 80-percent confidence intervals calculated from 10 equal-sized sets of samples drawn at random from the same waste stream

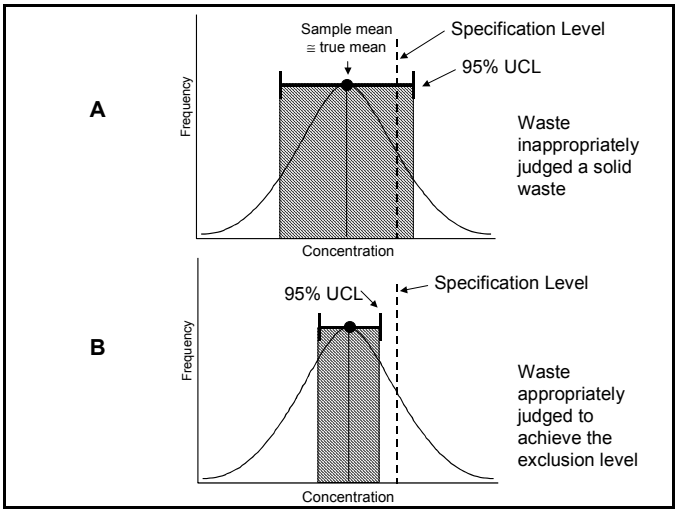

Figure 9. Example of how sampling precision could impact a waste exclusion demonstration under 40 CFR 261.38. Due to imprecision (A), the waste is inappropriately judged a solid waste. With more precise results (B), the entire confidence interval lies below the specification level, and the waste is appropriately judged eligible for the comparable fuels exclusion.

The *width* of the confidence interval (defined by the upper and lower confidence limits) is an indicator of the precision of the estimate of the parameter of interest. Generally, one can improve precision (i.e., reduce the standard error, $s / \sqrt{n}$) by taking more samples, increasing the physical size of each

sample (i.e., increasing the sample support), and by minimizing random variability introduced in the sampling and measurement processes.

For example, Figure 9 shows how sampling precision can affect the ability to claim an exclusion from the definition of solid waste under the comparable fuels regulations at 40 CFR 261.38. In Figure 9 "A," the sampling results are unbiased, but they are not sufficiently precise. In fact, the imprecision causes the confidence intervals to "straddle" the specification level; thus, there is not *statistically significant* evidence that the mean is below the standard. Imprecision can be caused by the heterogeneity of the material sampled, by random errors in the field and laboratory, and by too few samples. In Figure 9 "B," the results also are unbiased, but significant improvement in precision is observed (e.g., because more or larger samples were analyzed and errors were kept within acceptable limits), allowing us to conclude that the mean is indeed below the specification level.

Detailed guidance on the calculation of confidence limits for the mean can be found in Appendix F of this document.

3.4.2 Using a Proportion or Percentile To Determine Whether a Waste or Media Meets an Applicable Standard

Under RCRA, some regulatory thresholds are defined as concentration values that cannot be exceeded (e.g., the RCRA LDR program concentration-based treatment standards for hazardous waste specified at § 268.40 and § 268.48), concentration values that cannot be equaled or exceeded (e.g., the Toxicity Characteristic maximum concentration levels specified at § 261.24), or waste properties that cannot be exhibited (e.g., ignitability per § 261.21, corrosivity per § 261.22, or reactivity per § 261.23) for the waste to comply with the regulatory standard.

To demonstrate compliance with such a standard using sampling, it is necessary to consider the waste or site (whose boundaries are defined as a decision unit) as a population of discrete sample units (of a defined size, shape, and orientation). Ideally, none of these sample units may exceed the standard or exhibit the properties of concern for the waste or site to be in compliance with the standard. However, since it is not possible to know the status of all portions of a waste or site, samples must be used to infer - using statistical methods - what proportion or percentage of the waste complies, or does not comply, with the standard. Generally, few if any samples drawn from the population of interest may exceed the regulatory standard or exhibit the property of concern to demonstrate with reasonable confidence that a high proportion or percentage of the population complies with the standard.

Two simple methods for measuring whether a specified proportion or percentile of a waste or media meets an applicable standard are described in the following sections:

- Using an upper confidence limit on a percentile to classify a waste or media (Section 3.4.2.1), and
- Using a simple exceedance rule method to classify a waste or media (Section 3.4.2.2).

3.4.2.1 Using a Confidence Limit on a Percentile to Classify a Waste or Media

A percentile is a population parameter. We cannot know the true value of that parameter, but we can *estimate* it from a statistical sample drawn from the population by using a **confidence interval for a percentile**. If the **upper confidence limit (UCL) on the upper percentile** is below the fixed standard, then there is statistically significant evidence that the specified proportion of the waste or media attains the standard (see Figure 10). If the UCL on the upper percentile exceeds the standard (but all sample analysis results are below the standard), then the waste or media still could be judged in compliance with the standard; however, you would not have the specified degree of confidence that the specified proportion of the waste or media complies with the standard (see also the exceedance rule method, Section 3.4.2.2).

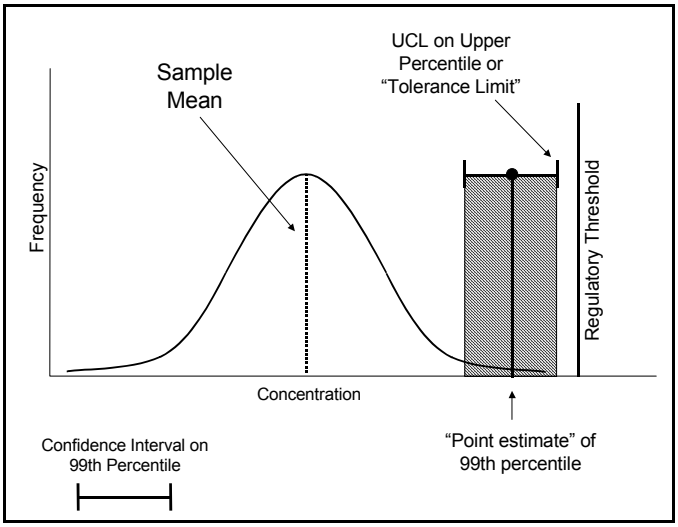

Figure 10. For a high percentile (e.g., the 99th percentile) to be less than an applicable standard, the mean concentration must be well below the standard.

Detailed guidance on the calculation of confidence limits for percentiles can be found in Section 8.2 and Appendix F of this document. Methods also are given in Conover (1999), Gilbert (1987, page 136), Hahn and Meeker (1991), and USEPA (1989a). A possible alternative to using a confidence limit on a percentile is the use of the "one-sample test for proportions" (see Section 3.2.2.1 of USEPA 2000d).

3.4.2.2 Using a Simple Exceedance Rule Method To Classify a Waste

One of the most straightforward methods for determining whether a given proportion or percentage of a waste (that is, all possible samples of a given sample support) complies with an applicable standard is to use a simple exceedance rule. To apply the method, simply obtain a number of samples and require that zero or few sample analysis results be allowed to exceed the applicable standard or possess the property (or "attribute") of interest. The method (also known as "inspection by attributes") is from a class of methods known as acceptance sampling plans (Schilling 1982, ASQ 1988 and 1993, and DoD 1996). One simple form of the exceedance rule, sometimes used by regulatory enforcement agencies, specifies zero exceedances in a set of samples. This method can be used to classify a waste (i.e., determine if it exhibits the characteristics of ignitability, corrosivity, reactivity[1], or toxicity) or to determine its status under RCRA (that is, to determine if the waste is prohibited from land disposal or if it attains an LDR treatment standard).

The method is attractive because it is simple (e.g., because sample analysis results are

[1] EPA uses a narrative criteria to define most reactive wastes, and waste handlers should use their knowledge to determine if a waste is sufficiently reactive to be regulated.

recorded as either “pass” or “fail” and statistical tables can be used instead of equations), it does not require an assumption about the form of the underlying distribution, and it can be used when a large proportion of the data are reported as less than a quantitation limit. Furthermore, the method has statistical properties that allow the waste handler to have a known level of confidence that at least a given proportion of the waste complies with the standard. One potential drawback of using an exceedance rule is that with a small number of samples, you might not be able to conclude with high confidence that a high proportion of the waste complies with the applicable standard (unless you have sufficient knowledge of the waste indicating there is little variability in concentrations or properties). That is, with a small number of samples, there is little statistical power: an unacceptably large proportion of the waste or site could exceed the standard or exhibit the property even though no such exceedances or properties were observed in the samples. Increasing the number of samples will improve the statistical performance.

As a practical matter, it is suggested that you scale the statistical performance and acceptance requirements (and thus, the number of samples) to the size of the lot or batch of waste of interest. For example, when large and/or very heterogeneous volumes of waste are the subject of the study, decision-makers may require high confidence that a high proportion of the waste meets the applicable standard. A relatively large number of samples will be required to satisfy these criteria if the exceedance rule is used. On the other hand, decision-makers may choose to relax the statistical performance criteria when characterizing a small volume of waste (or a very homogeneous waste) and thus fewer samples would be needed.

Detailed guidance on the use of an exceedance rule is provided in Section 5.5.2 and in Appendix F, Section F.3.2, of this document. The exceedance rule method also is described in *Methods for Evaluating the Attainment of Cleanup Standards. Volume 1: Soils and Solid Media* (USEPA 1989a, Section 7.4).

3.4.3 Comparing Two Populations

Some environmental studies do not involve testing compliance against a fixed standard but require comparison of two separate data. This type of analysis is common for detecting releases to ground water at waste management units such as landfills and surface impoundments, detecting releases to soil and the unsaturated zone at land treatment units, or determining if site contamination is distinguishable from natural background concentrations. In these situations, the operator must compare “on site” or “downgradient” concentrations to “background.”

For example, at a new land-based waste management unit (such as a new landfill), we expect the concentrations in a set of samples from downgradient locations to be similar to a set of samples from background locations. If a statistically significant change in downgradient conditions is detected, then there may be evidence of a release to the environment. Statistical methods called *two-sample tests* can be used to make such comparisons (they are called two-sample tests because two *sets* of samples are used). A two-sample test also could be used to measure changes in constituent concentrations in a waste or soil “before” treatment and “after” treatment to assess the effectiveness of the treatment process (see USEPA 2002a).

For detailed guidance on the use of two-sample tests, see EPA’s G-9 guidance (USEPA 2000d) and EPA’s guidance on the statistical analysis of ground-water monitoring data (USEPA 1989b

and 1992b).

Note that detecting a release to the environment may not necessarily involve use of a statistical test and may not even involve sampling. For example, observation of a broken dike at a surface impoundment may indicate that a release has occurred.

3.4.4 Estimating Spatial Patterns

Under some circumstances, a site investigator may wish to determine the location of a contaminant in the environment as well as its concentration. Knowledge of spatial trends or patterns may be of particular value when conducting risk assessments or locating areas for clean-up or removal under the RCRA Corrective Action program. Estimation of spatial patterns is best addressed by geostatistics or other spatial data analysis methods.

Geostatistical models are based on the notion that elements of the population that are close together in space and/or time exhibit an identifiable relationship or positive correlation with one another. Geostatistical techniques attempt to recognize and describe the pattern of spatial dependence and then account for this pattern when generating statistical estimates. On the other hand, “classical” methods assume that members of a population are not correlated (USEPA 1997a).

While a full treatment of spatial analysis and geostatistics is beyond the scope of this guidance, certain techniques recommended in the guidance require consideration of spatial differences. For example, you may need to consider whether there are any spatial correlations in a waste or site when selecting a sampling design. There are some relatively simple graphical techniques that can be used to explore possible spatial patterns or relationships in data. For example, posting plots or spatial contour maps can be generated manually or via software (e.g., see EPA’s Geo-EAS software described in Appendix H). Interested readers can find a more comprehensive explanation of spatial statistics in texts such as Myers (1997), Isaaks and Srivastava (1989), Journel (1988), USEPA (1991a, 1997a), or consult a professional environmental statistician or geostatistician.

4 PLANNING YOUR PROJECT USING THE DQO PROCESS

To be successful, a waste-testing program must yield data of the type and quality necessary to achieve the particular purpose of the program. This is accomplished through correct, focused, and well-documented sampling, testing, and data evaluation activities. In each case, a clear understanding of the program objectives and thorough planning of the effort are essential for a successful, cost-effective waste-testing program.

Each program design is unique because of the many possible variables in waste sampling and analysis such as regulatory requirements, waste and facility-specific characteristics, and objectives for the type and quantity of data to be provided. Nonetheless, a systematic planning process such as the Data Quality Objectives (DQO) Process, which takes these variables into account, can be used to guide planning efforts. EPA recommends using the DQO Process when data are being used to select between two opposing conditions, such as determining compliance with a standard.

The DQO Process yields qualitative and quantitative statements that:

- Clarify the study objectives
- Define the type, quantity, and quality of required data
- Determine the most appropriate conditions from which to collect the samples
- Specify the amount of uncertainty you are willing to accept in the results
- Specify how the data will be used to test a decision rule.

The outputs of the DQO Process are used to define the quality control requirements for sampling, analysis, and data assessment. These requirements are then incorporated into a QAPP, WAP, or other similar planning document.

The DQO Process comprises seven planning steps depicted in Figure 11. The figure shows one of the most important features of the process: its iterative nature. You don't have to "get it right the first time." You can use existing information to establish DQOs. If the initial design is not feasible, then you can iterate through one or more of the earlier planning steps to identify a sampling design that will meet the budget and generate data that are adequate for the decision. This way, you can evaluate sampling designs and related costs *in advance* before significant time and resources are expended to collect and analyze samples.

In a practical sense, the DQO Process offers a structured approach to "begin with the end in

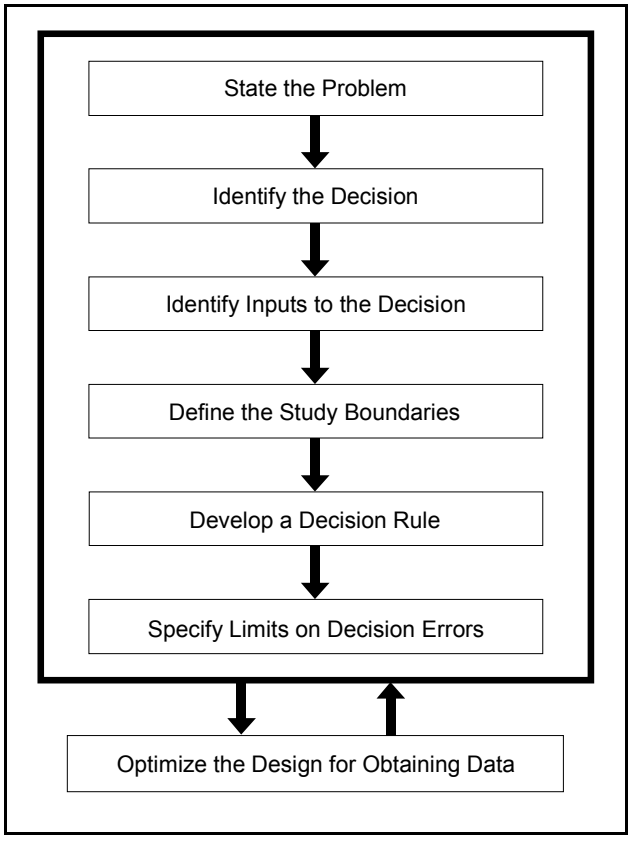

Figure 11. The seven steps of the DQO Process (from USEPA 2000b)

mind." It is a framework for asking the right questions and using the answers to develop and implement a cost-effective plan for data collection. The DQO Process does not necessarily proceed in a linear fashion or involve rigid procedures; rather, it is a thought process to enable you to get useful information in a cost-effective manner.

Systematic Planning and the DQO Process: EPA References and Software

Guidance for the Data Quality Objectives Process, EPA QA/G-4, August 2000, EPA/600/R-96/055. Provides guidance on how to perform the DQO Process.

Data Quality Objectives Decision Error Feasibility Trials Software (DEFT) - User's Guide, EPA QA/G-4D, September 2001, EPA/240/B-01/007 (User's Guide and Software). PC-based software for determining the feasibility of data quality objectives defined using the DQO Process.

Guidance for the Data Quality Objectives Process for Hazardous Waste Sites, EPA QA/G-4HW, January 2000, EPA/600/R-00/007. Provides guidance on applying the DQO Process to hazardous waste site investigations.

Failure to establish DQOs before implementing field and laboratory activities can cause difficulties in the form of inefficiencies, increased or unnecessary costs, or the generation of unusable data. For example, if the limit of quantitation for sample analysis is greater than the Action Level, then the data will not be useable for its intended purpose; or, if you do not collect enough samples, then you may not be able to draw conclusions with the desired level of confidence.

When properly used, the DQO Process:

- Provides a good way to document the key activities and decisions necessary to address the problem and to communicate the approach to others.
- Involves key decision makers, other data users, and technical experts in the planning process before data collection begins which helps lead to a consensus prior to beginning the project and makes it easier to change plans when circumstances warrant because involved parties share common understandings, goals, and objectives.
- Develops a consensus approach to limiting decision errors that strikes a balance between the cost of an incorrect decision and the cost of reducing or eliminating the possible mistake.
- Saves money by greatly reducing the tendency to collect unneeded data by encouraging the decision makers to focus on data that support only the decision(s) necessary to solve the problem(s). When used with a broader perspective in mind, however, the DQO Process may help identify opportunities to consolidate multiple tasks and improve the efficiency of the data collection effort.[1]

[1] In some cases, it might be appropriate and cost-effective to collect data beyond that required to support a near-term decision. For example, if a drill rig is mobilized to collect deep soil samples to determine the need for remediation, it would be cost-effective to also collect relatively low-cost data (such as geotechnical parameters, total organic carbon, moisture content, etc.) needed by engineers to design the remedy. Otherwise, unnecessary costs might be incurred to remobilize a drill rig to obtain data that could have been obtained in the initial effort.

The remainder of this section addresses how the DQO Process can be applied to RCRA waste-characterization studies. While the discussion is based on EPA's *G-4* guidance (USEPA 2000b), some steps have been modified or simplified to allow for flexibility in their use. Keep in mind that not all projects or decisions (such as a hazardous waste determination) will require the full level of activities described in this section, but the logic applies nonetheless. In fact, EPA encourages use of a "graded approach" to quality assurance. A graded approach bases the level of management and QA/QC activities on the intended use of the results and the degree of confidence needed in their quality (USEPA 2001f).

4.1 Step 1: State the Problem

Before developing a data gathering program, the first step is to state the problem or determine what question or questions are to be answered by the study. For many waste characterization or monitoring programs the questions are spelled out in the applicable regulations; however, in some cases, determining the actual problem or question to be answered may be more complex. As part of this step, perform the four activities described in the following sections.

DQO Step 1: State the Problem

Purpose
To define the problem so that the focus of the study will be unambiguous.

Activities
- Identify members of the planning team.
- Identify the primary decision maker(s).
- Develop a concise description of the problem.
- Determine resources – budget, personnel, and schedule.

4.1.1 Identify Members of the Planning Team

The planning team comprises personnel representing all phases of the project and may include stakeholders, decision makers, technical project managers, samplers, chemists, process engineers, QA/QC managers, statisticians, risk assessors, community leaders, grass roots organizations, and other data users.

4.1.2 Identify the Primary Decision Maker

Identify the primary decision maker(s) or state the process by which the decision will be made (for example, by consensus).

4.1.3 Develop a Concise Description of the Problem

Develop a problem description to provide background information on the fundamental issue to be addressed by the study. For RCRA waste-related studies, the "problem" could involve determining one of the following: (1) if a solid waste should be classified as a hazardous waste, (2) if a hazardous waste is prohibited from land disposal, (3) if a treated hazardous waste attains the applicable treatment standard, (4) if a cleanup goal has been attained, or (5) if hazardous constituents have migrated from a waste management unit.

Summarize existing information into a "conceptual model" or conceptual site model (CSM) including previous sampling information, preliminary estimates of summary statistics such as the mean and standard deviation, process descriptions and materials used, and any spatial and temporal boundaries of the waste or study area that can be defined. A CSM is a

three-dimensional “picture” of site conditions at a discrete point in time (a snapshot) that conveys what is known or suspected about the facility, releases, release mechanisms, contaminant fate and transport, exposure pathways, potential receptors, and risks. The CSM does not have to be based on a mathematical or computer model, although these tools often help to visualize current information and predict future conditions. The CSM should be documented by written descriptions of site conditions and supported by maps, cross sections, analytical data, site diagrams that illustrate actual or potential receptors, and any other descriptive, graphical, or tabular illustrations necessary to present site conditions.

4.1.4 Specify Available Resources and Relevant Deadlines

Identify available financial and human resources, identify deadlines established by permits or regulations, and establish a schedule. Allow time for developing acceptance and performance criteria, preparing planning documents (such as a QAPP, sampling plan, and/or WAP), collecting and analyzing samples, and interpreting and reporting data.

4.2 Step 2: Identify the Decision

The goal of this step is to define the questions that the study will attempt to answer and identify what actions may be taken based on the outcome of the study. As part of this step, perform the four activities described in the following sections.

DQO Step 2: Identify the Decision

Purpose
To define what specific decisions need to be made or what questions need to be answered.

Activities
- Identify the principal study question.
- Define the alternative actions that could result from resolution of the principal study question.
- Develop a decision statement.
- Organize multiple decisions.

4.2.1 Identify the Principal Study Question

Based on the problem identified in Step 1, identify the study question and state it as specifically as possible. This is an important step because the manner in which you frame the study question can influence whether sampling is even appropriate, and if so, how you will evaluate the results. Here are some examples of study questions that might be posed in a RCRA-related waste study:

- Does the filter cake from the filter press exhibit the TC at its point of generation?
- Does the treated waste meet the universal treatment standard (UTS) for land disposal under 40 CFR 268?
- Has the soil remediation at the SWMU attained the cleanup goal for benzene?
- Have hazardous constituents migrated from the land treatment unit to the underlying soil at concentrations significantly greater than background concentrations?
- Are radioactive and hazardous wastes colocated, producing a mixed waste management scenario?

Before conducting a waste-sampling and testing program to comply with RCRA, you should review the specific regulatory requirements in 40 CFR in detail and consult with staff from your EPA region or the representative from your State (if your State is authorized to implement the regulation).

4.2.2 Define the Alternative Actions That Could Result from Resolution of the Principal Study Question

Generally, two courses of action will result from the outcome of the study. One that involves action, such as deciding to classify a solid waste as a hazardous waste, and one that requires an alternative action, such as deciding to classify a solid waste as a nonhazardous solid waste.[2]

4.2.3 Develop a Decision Statement

In performing this activity, simply combine the principal study question and the alternative actions into a "decision statement." For example, you may wish to determine whether a waste exhibits a hazardous waste characteristic. The decision statement should be in writing (for example, in the QAPP) and agreed upon by the planning team. This approach will help avoid misunderstandings later in the process.

4.2.4 Organize Multiple Decisions

If several separate decisions statements must be defined to address the problem, then you should list them and identify the sequence in which they should be resolved. For example, if you classify a solid waste as a nonhazardous waste, then you will need to make a waste management decision. Options might include land disposal (e.g., in an industrial landfill or a municipal solid waste landfill), recycling, or some other use. You might find it helpful to document the decision resolution sequence and relationships in a diagram or flowchart.

4.3 Step 3: Identify Inputs to the Decision

In most cases, it will be necessary to collect data or new information to resolve the decision statement. To identify the type and source of this information, perform the activities outlined in the following four sections.

DQO Step 3: Identify Inputs to the Decision

Purpose
To identify data or other information required to resolve the decision statement.

Activities
- Identify the information required to resolve the decision statement.
- Determine the sources of information.
- Identify information needed to establish the Action Level.
- Identify sampling and analysis methods that can meet the data requirements.

4.3.1 Identify the Information Required

For RCRA-related waste studies, information requirements typically will

[2] Testing alone might not be sufficient to determine if a solid waste is hazardous waste. You also should apply knowledge of the waste generation process to determine if the solid waste is a hazardous waste under 40 CFR 261.

include samples to be collected, variables to be measured (such as total concentrations, TCLP results, or results of tests for other characteristics, such as reactivity, ignitability, and corrosivity), the units of measure (such as mg/L), the form of the data (such as on a dry weight basis), and waste generation or process knowledge.

4.3.2 Determine the Sources of Information

Identify and list the sources of information needed and qualitatively evaluate the usefulness of the data. Existing information, such as analytical data, can be very valuable. It can help you calculate the appropriate number of new samples needed (if any) and reduce the need to collect new data (see also Section 5.4).

4.3.3 Identify Information Needed To Establish the Action Level

The Action Level is the threshold value that provides the criterion for choosing between alternative actions. Under RCRA, there are several types of Action Levels.

The first type of Action Level is a fixed standard or regulatory threshold (RT) usually specified as a *concentration* of a hazardous constituent (e.g., in mg/L). Examples of regulatory thresholds that are Action Levels in the RCRA regulations include the TC Regulatory Levels at 40 CFR 261.24 and the Land Disposal Restrictions (LDR) numeric treatment standards at 40 CFR 268.40.

Another criterion for choosing between alternative actions is defined by the *property* of a waste. Three such properties are defined in the RCRA regulations: ignitability (§ 261.21), corrosivity (§ 261.22), and reactivity (§ 261.23). The results of test methods used to determine if a waste is ignitable, corrosive, or reactive are interpreted as either “pass” or “fail” -- i.e., the waste either has the property or it does not. Note that a concentration measurement, such as a TCLP sample analysis result, also can be interpreted as either “pass” or “fail” based on whether the value is less than or greater than a specified threshold.

A third criterion for choosing between alternative actions involves making a comparison between constituent concentrations at different times or locations to determine if there has been a change in process or environmental conditions over time. In these situations, you need to determine if the two sets of data are different relative to each other rather than checking for compliance with a fixed standard.

Finally, an Action Level can represent a proportion of the population having (or not having) some characteristic. For example, while it might be desirable to have all portions of a waste or site comply with a standard, it would be more practical to test whether some high proportion (e.g., 0.95) of units of a given size, shape, and orientation comply with the standard. In such a case, the Action Level could be set at 0.95.

For more information on identifying the Action Level, see Section 2 (RCRA regulatory drivers for waste sampling and testing), the RCRA regulations in 40 CFR, ASTM Standard D 6250 (*Standard Practice for Derivation of Decision Point and Confidence Limit for Statistical Testing of Mean Concentration in Waste Management Decisions*), or consult with your State or EPA Regional staff.

4.3.4 Confirm That Sampling and Analytical Methods Exist That Can Provide the Required Environmental Measurements

Identify and evaluate candidate sampling and analytical methods capable of yielding the required environmental measurements. You will need to revisit this step during Step 7 of the DQO Process ("Optimize the Design for Obtaining the Data") after the quantity and quality of the necessary data are fully defined. In evaluating sampling methods, consider the medium to be sampled and analyzed, the location of the sampling points, and the size, shape and orientation of each sample (see also Section 6, "Controlling Variability and Bias in Sampling" and Section 7, "Implementation: Selecting Equipment and Conducting Sampling").

In evaluating analytical methods, choose the appropriate candidate methods for sample analyses based on the sample matrix and the analytes to be determined.

Guidance on the selection of analytical methods can be found in Chapter Two of SW-846 ("Choosing the Correct Procedure"). Up-to-date information on analytical methods can be found at SW-846 "On Line" at http://www.epa.gov/epaoswer/hazwaste/test/main.htm.

4.4 Step 4: Define the Study Boundaries

In this step of the DQO Process, you should identify the target population of interest and specify the spatial and temporal features of that population that are pertinent for decision making.

To define the study boundaries, perform the activities described in the following five sections.

DQO Step 4: Define the Study Boundaries

Purpose
To define the spatial and temporal boundaries that are covered by the decision statement.

Activities
- Define the target population of interest.
- Define the "sample support"
- Define the spatial boundaries that clarify what the data must represent.
- Define the time frame for collecting data and making the decision.
- Identify any practical constraints on data collection.
- Determine the smallest subpopulation, area, volume, or time for which separate decisions must be made.

4.4.1 Define the Target Population of Interest

It is important for you to clearly define the target population to be sampled. Ideally, the target population coincides with the population to be sampled (Cochran 1977) – that is, the target population should represent the total collection of all possible sampling units that could be drawn. Note that the "units" that make up the population are defined operationally based on their size, shape, orientation, and handling (i.e., the "sample support").[3] The sampling unit definition must be considered when defining the target population because any changes in the definition can affect the population characteristics. See Section 6.3.1 for guidance on establishing the appropriate size (mass) of a sample, and see Section 6.3.2 for guidance on

[3] The physical size (expressed as mass or volume), shape, and orientation of a sample is known as the *sample support*. Sample support plays an important role in characterizing waste or environmental media and in minimizing variability caused by the sampling process. The concept of *support* is discussed in greater detail in Section 6.2.3.

establishing the appropriate shape and orientation of sample.

Define the target population in terms of sampling units, the decision-making volume, and the location of that volume.

Sampling at the **point of generation** is *required* by regulation when determining the regulatory status of a waste. See 55 FR 11804, March 29, 1990, and 55 FR 22652, June 1, 1990.

4.4.2 Define the Spatial Boundaries

If sampling at the point of waste generation (i.e., *before* the waste is placed in a container or transport unit), then the sampling problem could involve collecting samples of a moving stream of material, such as from a conveyor, discharge pipe, or as poured into a container or tank. If so, then physical features such as the width of the flow or discharge and the rate of flow or discharge will be of interest for defining the spatial boundary of the problem.

If the sampling problem involves collecting samples from a waste storage unit or transport container, then the spatial boundaries can be defined by some physical feature, such as volume, length, width, height, etc. The spatial boundaries of most waste storage units or containers can be defined easily. Examples of these units follow:

- Container such as a drum or a roll-off box
- Tank
- Surface Impoundment
- Staging Pile
- Waste Pile
- Containment Building.

In other cases, the spatial boundary could be one or more geographic areas, such as areas representing “background” and “downgradient” conditions at a land treatment unit. Another example is a SWMU area that has been subject to remediation where the objective is verify that the cleanup goal has been achieved over a specified area or volume at the SWMU. If the study requires characterization of subsurface soils and ground water, then consult other guidance (for example, see USEPA 1989a, 1989b, 1991d, 1992a, 1993c, and 1996b).

To help the planning team visualize the boundary, it may be helpful to prepare a drawing, map, or other graphical image of the spatial boundaries, including a scale and orientation (e.g., a north arrow). If appropriate and consistent with the intended use of the information, maps also should identify relevant surface features (such as buildings, structures, surface water bodies, topography, etc.) and known subsurface features (pipes, utilities, wells, etc.).

If samples of waste will be taken at the point of generation (e.g., when the waste becomes a solid waste), the location of that point should be defined in this step of the DQO Process.

4.4.3 Define the Temporal Boundary of the Problem

A temporal boundary could be defined by a permit or regulation (such as the waste generated per day) or operationally (such as the waste generated per “batch” or truck load). You should

determine the time frame to which the decision applies and when to collect the data. In some cases, different time intervals might be established to represent different populations (e.g., in the case where there is a process change over time that affects the character of the waste).

Waste characteristics or chemistry, such as the presence of volatile constituents, also could influence the time frame within which samples are collected. For example, volatilization could occur over time.

4.4.4 Identify Any Practical Constraints on Data Collection

Identify any constraints or obstacles that could potentially interfere with the full implementation of the data collection design. Examples of practical constraints include physical access to a sampling location, unfavorable weather conditions, worker health and safety concerns, limitations of available sampling devices, and availability of the waste (e.g., as might be the case for wastes generated from batch processes) that could affect the schedule or timing of sample collection.

4.4.5 Define the Scale of Decision Making

Define the smallest, most appropriate subsets of the population (sub-populations), waste, or media to be characterized based on spatial or temporal boundaries. The boundaries will define the unit of waste or media about which a decision will be made. The unit is known as the **decision unit**.

When defining the decision unit, the consequences of making a decision error should be carefully considered. The consequences of making incorrect decisions (Step 6) are associated with the size, location, and shape of the decision unit. For example, if a decision, based on the data collected, results in a large volume of waste being classified as nonhazardous, when in fact a portion of the waste exhibits a hazardous waste characteristic (e.g., due to the presence of a "hot spot"), then the waste generator could potentially be found in violation of RCRA . To limit risk of managing hazardous waste with nonhazardous waste, the waste handler should consider dividing the waste stream into smaller decision units – such as the volume of waste that would be placed into an individual container to be shipped for disposal – and make a separate waste classification decision regarding each decision unit.

The planning team may establish decision units based on several considerations:

- **Risk** – The scale of the decision making could be defined based on an exposure scenario. For example, if the objective is to evaluate exposures via direct contact with surface soil, each decision unit could be defined based on the geographic area over which an individual is assumed to move randomly across over time. In EPA's Superfund program, such a unit is known as an "exposure area" or EA (USEPA 1992c and 1996f). An example of an EA from EPA's *Soil Screening Guidance: User's Guide* (USEPA 1996f) is the top 2 centimeters of soil across a 0.5-acre area. In this example, the EA is the size of a suburban residential lot and the depth represents soil of the greatest concern for incidental ingestion of soil, dermal contact, and inhalation of fugitive dust.

  If evaluation of a decision unit or EA for the purpose of making a cleanup

decision finds that cleanup is needed, then the same decision unit or EA should be used when evaluating whether the cleanup standard has been attained. Furthermore, the size, shape, and orientation (the "sample support") of the samples used to determine that cleanup was necessary should be the same for samples used to determine whether the cleanup standard is met (though this last condition is not strictly necessary when the parameter of interest is the mean).

- **Operational Considerations** – The scale of the decision unit could be defined based on operational considerations, such as the need to characterize each "batch" of waste after it has been treated or the need to characterize each drum as it is being filled at the point of waste generation. As a practical matter, the scale for the decision making often is defined by the spatial boundaries – for example as defined by a container such as a drum, roll-off box, truck load, etc. or the time required to fill the container.

- **Other** – The possibility of "hot spots" (areas of high concentration of a contaminant) may be apparent to the planning team from the history of the facility. In cases where previous knowledge (or planning team judgment) includes identification of areas that have a higher potential for contamination, a scale may be developed to specifically represent these areas.

Additional information and considerations on defining the scale of the decision making can be found in *Guidance for the Data Quality Objectives Process for Hazardous Waste Site Operations EPA QA/G-4HW* (USEPA 2000a) and *Guidance for the Data Quality Objectives Process EPA QA/G-4* (USEPA 2000b).

4.5 Step 5: Develop a Decision Rule

A statement must be developed that combines the parameter of interest and the Action Levels with the DQO outputs already developed. The combination of these three elements forms the decision rule and summarizes what attributes the decision maker wants to study and how the information will assist in solving the central problem. To develop the decision rule, perform the activities described in the following three sections:

4.5.1 Specify the Parameter of Interest

DQO Step 5: Develop a Decision Rule

Purpose
To define the parameter of interest, specify the Action Level and integrate previous DQO outputs into a single statement that describes a logical basis for choosing among alternative actions; i.e., define how the data will be used to make a decision.

Activities
- Specify the parameter of interest (mean, median, percentile).
- Specify the Action Level for the study.
- Develop a decision rule.

A statistical "parameter" is a descriptive measure of a population such as the population mean, median, or a percentile (see also Section 3.2). See Table 2.

Some of the RCRA regulations specify the parameter of interest. For example, the comparable fuels sampling and analysis requirements at 40 CFR 261.38(c)(8)(iii)(A) specify the *mean* as the parameter of interest, and the ground-water monitoring requirements at 40 CFR 264.97 specify the parameter of interest for each statistical

test. Other RCRA regulations do not specify the parameter of interest, however, you can select a parameter based on what the Action Level is intended to represent. In general, if an Action Level is based on long-term average health effects, the parameter of interest could be the population mean (USEPA 1992a). If the Action Level represents a value that should never (or rarely) be exceeded, then the parameter of interest could be an upper population percentile, which can serve as a reasonable approximation of the *maximum* value.

If the objective of the study does not involve estimation of a parameter or testing a hypothesis, then specification of a parameter is not necessary.

Table 2. Population Parameters and Their Applicability to a Decision Rule

Parameter	***Definition***	***Appropriate Conditions for Use***
Mean	Average	Estimate central tendency: Comparison of middle part of population to an Action Level.
Median	Middle observation of the distribution; 50th percentile; half of data are above and below	May be preferred to estimate central tendency if the population contains many values that are less than the limit of quantitation. The median is not a good choice if more than 50% of the population is less than the limit of quantitation because a true median does not exist in this case. The median is not influenced by the extremes of the contaminant distribution.
Percentile	Specified percent of sample that is equal to or below the given value	For cases where it is necessary to demonstrate that, at most, only a small portion of a population could exceed the Action Level. Sometimes selected if the decision rule is being developed for a chemical that can cause acute health effects. Also useful when a large part of the population contains values less than the detection limit.

4.5.2 Specify the Action Level for the Study

You should specify an Action Level or concentration limit that would cause the decision maker to choose between alternative actions. Examples of Action Levels follow:

- Comparable/syngas fuel constituent specification levels specified at § 261.38
- Land disposal restrictions concentration level treatment standards at § 268.40 and § 268.48
- Risk-based cleanup levels specified in a permit as part of a corrective action
- "Pass" or "fail" thresholds for tests for ignitability, corrosivity, reactivity[4], and toxicity.

Also, be sure the detection or quantitation limits for the analytical methods identified in DQO Step 3 (Section 4.3) are below the Action Level, if possible.

[4] EPA uses a narrative criteria to define most reactive wastes, and waste handlers should use their knowledge to determine if a waste is sufficiently reactive to be regulated.

If your objective is to compare "onsite" to "background" to determine if there is a statistically significant increase above background (as would be the case for monitoring releases from a land treatment unit under § 264.278), you will not need to specify an Action Level; rather, the Action Level is implicitly defined by the background concentration levels and the variability in the data. A summary of methods for determining background concentrations in soil can be found in USEPA 1995a. Methods for determining background concentrations in ground water can be found in USEPA 1989b and 1992b.

Finally, note that some studies will not require specification of a regulatory or risk-based Action Level. For example, if the objective may be to identify the existence of a release, samples could be obtained to verify the *presence or absence* of a spill, leak, or other discharge to the environment. Identifying a potential release also could include observation of abandoned or discarded barrels, containers, and other closed receptacles containing hazardous wastes or constituents (see 61 FR No. 85, page 19442).

4.5.3 Develop a Decision Rule

After you have completed the above activities, you can construct a decision rule by combining the selected population parameter and the Action Level with the scale of the decision making (from DQO Process Step 4) and the alternative action (from DQO Step 2). Decision rules are expressed as "if (criterion)..., then (action)...." A hypothetical example follows:

> "If the true 95th percentile of all possible 100-gram samples of the waste being placed in the 20-cubic yard container is less than 5.0 mg/L TCLP lead, then the solid waste will be classified as nonhazardous waste. Otherwise, the solid waste will be classified as a RCRA hazardous waste."

Note that this is a functional decision rule based on an ideal condition (i.e., knowledge of the true concentration that equals the 95th percentile of all possible sample analysis results). It also identifies the boundary of the study by specifying the sample unit (100-gram samples in accordance with the TCLP) and the size of the decision unit. It does *not*, however, specify the amount of uncertainty the decision maker is willing to accept in the estimate. You specify that in the next step.

4.6 Step 6: Specify Limits on Decision Errors

Step 6: Specify Limits on Decision Errors

Purpose
To specify the decision maker's tolerable limits on decision error.

Activities
- Identify potential sources of variability and bias in the sampling and measurement processes (see Section 6)
- Determine the possible range on the parameter of interest.
- Choose the null hypothesis.
- Consider the consequences of making an incorrect decision.
- Specify a range of values where the consequences are minor (the "gray region")
- Specify an acceptable probability of making a decision error.

Because samples represent only a portion of the population, the information available to make decisions will be incomplete; hence, *decision errors* sometimes will be made. Decision errors occur because decisions are made using *estimates* of the parameter of interest, rather than the true (and unknown) value. In fact, if you repeatedly sampled and analyzed a waste over and over in an identical manner the results would be a little different each time (see Figure 8 in Section 3). This variability

in the results is caused by the non-homogeneity of the waste or media, slight differences in how the samples of the waste were collected and handled, variability in the analysis process, and the fact that only a small portion of the waste is usually ever sampled and tested. (See Section 6.1 for a more detailed discussion of sources of variability and bias in sampling). For example, if you conduct sampling and analysis of a solid waste and classify it as "nonhazardous" based on the results, when in fact it *is* a hazardous waste, you will have made a wrong decision or *decision error*. Alternatively, if you classify a solid waste as hazardous, when in fact it is nonhazardous, you also will have made a *decision error*.

There are two types of decision error. A "Type I" or "false rejection" decision error occurs if you reject the null hypothesis when it is true. (The "null hypothesis" is simply the situation presumed to be true or the "working assumption".) A "Type II" or "false acceptance" decision error occurs if you accept the null hypothesis when it is false.[5]

Table 3 summarizes the four possible situations that might arise when a hypothesis is tested. The two possible true conditions correspond to the two columns of the table: the null hypothesis or "baseline assumption" is either true or the alternative is true. The two kinds of decisions are shown in the body of the table. Either you decide the baseline is true, or you decide the alternative is true. Associated with these two decisions are the two types of risk – the risk of making a Type I (false rejection) error (denoted by α) and the risk of making a Type II (false acceptance) error (denoted by β). You can improve your chances of making correct decisions by reducing α and β (which often requires more samples or a different sampling design) and by using field sampling techniques that minimize errors related to sampling collection and handling (see also Sections 6 and 7).

Table 3. Conclusions and Consequences for a Test of Hypotheses

		True Condition	
		Baseline is True	**Alternative is True**
Decision Based on Sample Data	Baseline is True	Correct Decision	Type II (false acceptance) error (probability β)
	Alternative is True	Type I (false rejection) error (probability α)	Correct Decision

For many sampling situations under RCRA, the most conservative (i.e., protective of the environment) approach is to presume that the constituent concentration in the waste or media exceeds the standard in the absence of strong evidence to the contrary.[6] For example, in

[5] Statisticians sometimes refer to a Type I error as a "false positive," and a Type II error as a "false negative." The terms refer to decision errors made relative to a null hypothesis, and the terms may not necessarily have the same meaning as those used by chemists to describe analytical detection of a constituent when it is not really present ("false positive") or failure to detect a constituent when it really *is* present ("false negative").

[6] An exception to this assumption is found in "detection monitoring" and "compliance monitoring" in which underlying media (such as soil, pore water, or ground water) at a new waste management unit are presumed "clean" until a statistically significant increase above background is demonstrated (in the case of detection monitoring) or a statistically significant increase over a fixed standard is demonstrated (in the case of compliance or assessment monitoring).

testing a solid waste to determine if it exhibits the TC, the null hypothesis can be stated as follows: "the concentration is equal to or greater than the TC regulatory level." The alternative hypothesis is "the concentration is less than the TC regulatory level." After completion of the sampling and analysis phase, you conduct an assessment of the data. If your estimate of the parameter of interest is less than the threshold when the true value of the parameter exceeds the threshold, you will make a decision error (a Type I error). If the estimate of the parameter of interest is greater than the threshold when the true value is less than the threshold, you also will make an error (a Type II error) -- but one that has little potential adverse impacts to human health and the environment.

Note that during the planning phase and during sampling you will not know which kind of error you might make. Later, after a decision has been made, if you *rejected* the null hypothesis then you either made a Type I (false rejection) decision error or not; you could not have made a Type II (false acceptance) decision error. On the other hand, if you did not reject the null hypothesis, then you either made a Type II (false acceptance) error or not; you could not have made a Type I (false rejection) error. In either case, you will know which type of error you might have made and you will know the *probability* that the error was made.

In the RCRA program, EPA is concerned primarily with controlling errors having the most adverse consequences for human health and the environment. In the interest of protecting the environment and maintaining compliance with the regulations, there is an incentive on the part of the regulated entity to minimize the chance of a Type I decision error. The statistical methods recommended in this document emphasize controlling the Type I (false rejection) error rate and do not necessarily require specification of a Type II (false acceptance) error rate.

The question for the decision maker then becomes, what is the acceptable probability (or chance) of making a decision error? To answer this question, four activities are suggested. These activities are based on guidance found in *Guidance for the Data Quality Objectives Process QA/G-4* (USEPA 2000b) but have been tailored for more direct application to RCRA waste-related studies. The *Guidance for the Data Quality Objectives Process EPA QA/G-4* also provides detailed guidance on the use of a graphical construct called a Decision Performance Curve to represent the quality of a decision process.

4.6.1 Determine the Possible Range on the Parameter of Interest

Establish the possible range (maximum and minimum values) of the parameter of interest using data from a pilot study, existing data for a similar waste stream, or process knowledge (e.g., using a materials-balance approach). It is desirable, but not required, to have an estimate of the standard deviation as well.

4.6.2 Identify the Decision Errors and Choose the Null Hypothesis

Table 4 presents four examples of decision errors that could be made in a RCRA waste study. In the first three examples, the consequences of making a Type I error could include increased risk to human health and the environment or a potential enforcement action by a regulatory authority. The consequences of making a Type II error could include unnecessary financial and administrative resources required to manage the waste as hazardous (when, in fact, it is not) or continuing site cleanup activities when, in fact, the site is "clean."

Table 4. Examples of Possible Decision Errors in RCRA Waste Studies

Regulatory Requirement	*"Null Hypothesis" (baseline condition)*	*Possible Decision Errors*	
		Type I Error (α) "False Rejection"	*Type II Error (β) "False Acceptance"*
Example 1: Under 40 CFR 261.11, conduct sampling to determine if a solid waste is a hazardous waste by the TC.	The solid waste contains TC constituents at concentrations equal to or greater than their applicable regulatory levels (i.e., the solid waste is a hazardous waste).	Concluding the waste is not hazardous when, in fact, it is.	Deciding the waste is hazardous when, in fact, it is not.
Example 2: Under 40 CFR 268.7, conduct sampling and testing to certify that a hazardous waste has been treated so that concentrations of hazardous constituents meet the applicable LDR treatment standards.	The concentration of the hazardous constituents exceeds the treatment standard (i.e., the treatment standard has not been attained).	Concluding the treatment standard has been met when, in fact, it has not.	Concluding the treatment standard has not been met when, in fact, it has.
Example 3: Under 40 CFR 264.101 (and proposed Subpart S - Corrective Action at SWMUs), a permittee conducts testing to determine if a remediation at a SWMU has attained the risk-based cleanup standard specified in the permit.*	The mean concentration in the SWMU is greater than the risk-based cleanup standard (i.e., the site is contaminated).†	Concluding the site is "clean" when, in fact, it is contaminated.	Concluding the site is still contaminated when, in fact, it is "clean."
Example 4: Under 40 CFR 264.98(f), detection monitoring, monitor ground water at a regulated unit to determine if there is a statistically significant increase of contamination above background.	The level of contamination in each point of compliance well does not exceed background.	Concluding the contaminant concentration in a compliance well exceeds background when, in fact, it does not.	Concluding the contaminant concentration in a compliance well is similar to background when, in fact, it is higher.

* If the cleanup standard is based on "background" rather than a risk-based cleanup standard, then the hypotheses would be framed *in reverse* where the mean background and on-site concentrations are presumed equal unless there is strong evidence that the site concentrations are greater than background.
† A parameter other than the mean may be used to evaluate attainment of a cleanup standard (e.g., see USEPA 1989a).

In Example 4, however, the null hypothesis is framed *in reverse* of Examples 1 through 3. When conducting subsurface monitoring to detect contamination at a new unit (such as in detection monitoring in the RCRA ground-water monitoring program), the natural subsurface environment is presumed uncontaminated until statistically significant increases over the background concentrations are detected. Accordingly, the null hypothesis is framed such that the downgradient conditions are consistent with the background. In this case, EPA's emphasis on the protection of human health and the environment calls for minimizing the Type II error -- the mistake of judging downgradient concentrations the same as the background when, in fact,

they are higher. Detailed guidance on detection and compliance monitoring can be found in *RCRA Ground-Water Monitoring: Draft Technical Guidance* (USEPA 1992c) and EPA's guidance on the statistical analysis of ground-water monitoring data at RCRA facilities (USEPA 1989b and 1992b).

4.6.3 Specify a Range of Possible Parameter Values Where the Consequences of a False Acceptance Decision Error are Relatively Minor (Gray Region)

The "gray region" is one component of the quantitative decision performance criteria the planning team establishes during the DQO Process to limit impractical and infeasible sample sizes. The gray region is a range of possible parameter values near the action level where it is "too close to call." This gray area is where the sample data tend toward rejecting the baseline condition, but the evidence (data statistics) is not sufficient to be overwhelming. In essence, the gray region is an area where it will not be feasible to control the false acceptance decision error limits to low levels because the high costs of sampling and analysis outweigh the potential consequences of choosing the wrong course of action.

In statistical language, the gray region is called the "minimum detectable difference" and is often expressed as the Greek letter delta (Δ). This value is an essential part of the calculations for determining the number of samples that need to be collected so that the decision maker may have confidence in the decision made based on the data collected.

The first boundary of the gray region is the Action Level. The other boundary of the gray region is established by evaluating the consequences of a false acceptance decision error over the range of possible parameter values in which this error may occur. This boundary corresponds to the parameter value at which the consequences of a false acceptance decision error are significant enough to have to set a limit on the probability of this error occurring. The gray region (or "area of uncertainty") establishes the minimum distance from the Action Level where the decision maker would like to begin to control false acceptance decision errors.

In general, the narrower the gray region, the greater the number of samples needed to meet the criteria because the area of uncertainty has been reduced.

The quality of the decision process, including the boundaries of the gray region, can be depicted graphically using a Decision Performance Goal Diagram (DPGD). Detailed guidance on the construction and use of DPGDs is given in EPA DQO guidance documents (e.g., USEPA 2000a and 2000b) and in *Data Quality Objectives Decision Error Feasibility Trials Software (DEFT) - User's Guide* (USEPA 2001a). Figure 12(a) and Figure 12(b) show how some of the key outputs of Step 6 of the DQO Process are depicted in a DPGD when the parameter of interest is the mean (Figure 12(a)) and a percentile (Figure 12(b) .

The DPGD given in Figure 12(a) shows how the boundaries of the gray region are set when the null hypothesis is established as "the true mean concentration exceeds the standard." Notice that the planning team has set the action level at 5 ppm and the other boundary of the gray region at 4 ppm. This implies that when the mean calculated from the sample data is less than 4 ppm (and the planning assumptions regarding variability hold true), then the data will be considered to provide "overwhelming evidence" that the true mean (unknown, of course) is below the action level.

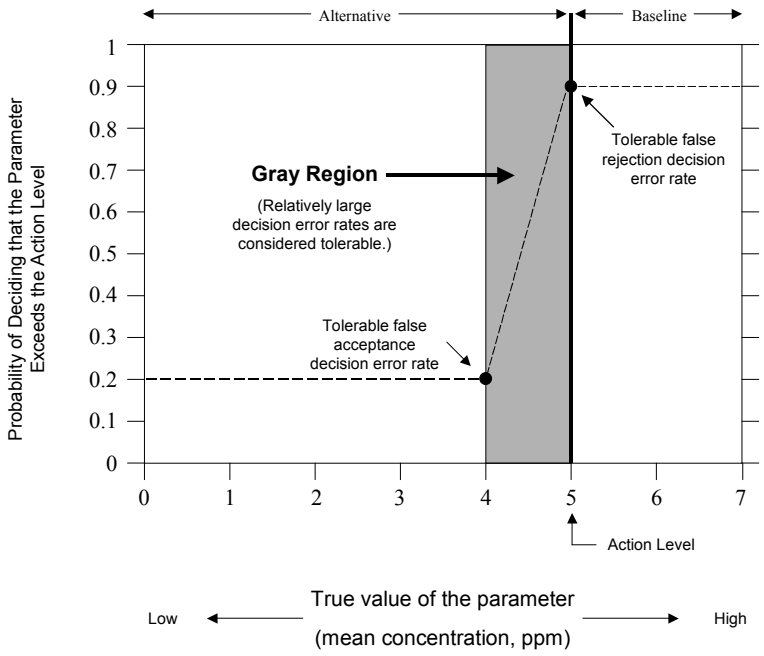

Figure 12(a). Decision Performance Goal Diagram where the mean is the parameter of interest. Null hypothesis (baseline condition): the true mean exceeds the action level.

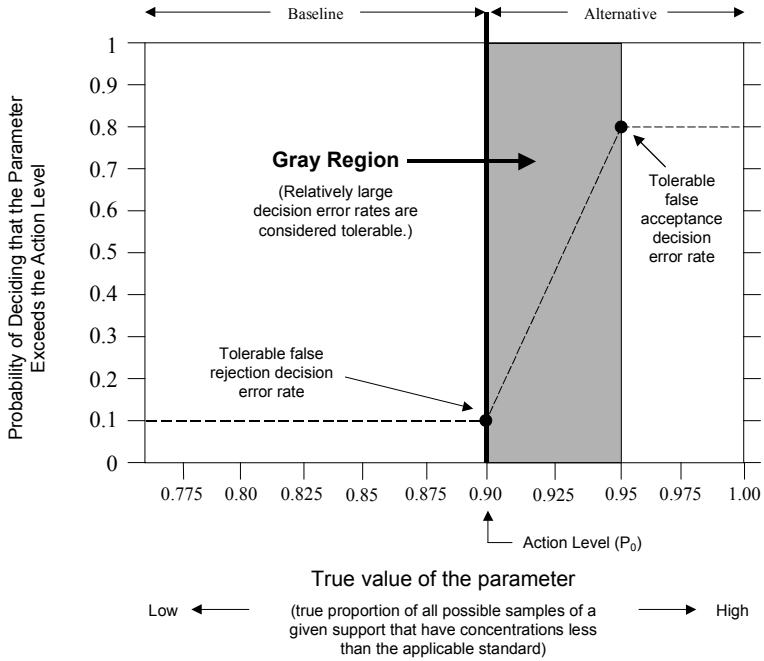

Figure 12(b). Decision Performance Goal Diagram where a percentile is the parameter of interest. Null hypothesis (baseline condition): true proportion -- of all possible samples of a given support that are less than the applicable standard -- is less than 0.90.

Now consider the DPGD given in Figure 12(b). The figure shows how the gray region is set when the null hypothesis is established as "the true proportion of samples below the concentration standard is less than 0.90." Notice in this example the planning team has set the action level at 0.90 and the other boundary of the gray region at 0.95. This implies that when the proportion of samples that comply with the standard is greater than 0.95, then the data will be considered to provide "overwhelming evidence" that the true proportion (unknown, of course) is greater than the action level of 0.90.

The term "samples" refers to all possible samples of a specified size, shape, and orientation (or **sample support**) drawn from the DQO decision unit. Sampling procedures and sample support can affect the measurement value obtained on individual samples and have a profound effect on the shape of the sampling distribution. Thus, the outcome of statistical procedures that examine characteristics of the upper tail of the distribution can be influenced by the sample support – more so than when the mean is the parameter of interest. Accordingly, when testing for a proportion, a complete statement of the null hypothesis should include specification of the sample support. See Sections 6.3.1 and 6.3.2 for guidance on establishing the appropriate sample support as part of the DQO Process.

4.6.4 Specify an Acceptable Probability of Making a Decision Error

You can never completely eliminate decision errors or even know when they have occurred, but you can quantify the probability of making such errors. In this activity, you establish the acceptable probability of making a decision error.

The Type I error rate (α) is a measure of the amount of "mistrust" you have in the conclusion (Myers 1997) and is also known as the **significance level** for a test. The flip side of this is the amount of faith or confidence you have in the conclusion. The **confidence level** is denoted mathematically as $1-\alpha$. As stated previously, the Type I error (the error of falsely rejecting the null hypothesis) is of greatest concern from the standpoint of environmental protection and regulatory compliance.

The probability of making a Type II error (the error of falsely accepting the null hypothesis) also can be specified. For example, if the sample data lead you to conclude that a waste does not qualify for the comparable fuels exclusion (40 CFR 261.38), when the true mean concentration in the waste is in fact below the applicable standard, then a Type II (false acceptance error) has been made. (Note that some of the statistical methods given in this document do not require specification of a Type II error rate).

As a general rule, the lower you set the probability of making a decision error, the greater the cost in terms of the number of samples required, time and personnel required for sampling and analysis, and financial resources required.

An acceptable probability level for making a decision error should be established by the planning team after consideration of the RCRA regulatory requirements, guidance from EPA or the implementing agency, the size (volume or weight) of the decision unit, and the consequences of making a decision error. In some cases, the RCRA regulations specify the Type I or Type II (or both) error rates that should be used. For example, when testing a waste to determine whether it qualifies for the comparable/syngas fuel exclusion under 40 CFR 261.38, the regulations *require* that the determination be made with a Type I error rate set at 5

percent (i.e., $\alpha = 0.05$).[7]

In other cases, the regulations do not specify any decision error limits. The planning team must specify the decision error limits based on their knowledge of the waste; impacts on costs, human health, and ecological conditions; and the potential consequences of making a decision error. For example, if the quantity of waste (that comprises a decision unit) is large and/or heterogeneous, then a waste handler may require high confidence (e.g., 95 or 99 percent) that a high proportion of the waste or media complies with the applicable standard. On the other hand, if the waste quantity is a relatively small (e.g., a drum) and sampling and measurement error can be minimized, then the waste handler may be willing to relax the confidence level required or simply use a nonstatistical (e.g., judgmental) sampling design and reduce the number of samples to be taken.

For additional guidance on controlling errors Section 6 and EPA's DQO guidance (USEPA 2000a and 2000b).

4.7 Outputs of the First Six Steps of the DQO Process

Table 5 provides a summary of the outputs of the first six steps of the DQO Process. Typically, this information will be incorporated into a QAPP, WAP, or other similar planning document (as described in Section 5.7). The DQOs can be simple and straight forward for simple projects and can be documented in just a few pages with little or no supporting data. For more complex projects, the DQOs can be more lengthy, and the supporting data may take up volumes. The team that will be optimizing the sample design(s) will need the information to support their plan development. The project manager and the individuals who assess the overall outcome of the project also will need the information to determine if the DQOs were achieved.

Keep in mind that the DQO Process is an iterative one; it might be necessary to return to earlier steps to modify inputs when new data become available or to change assumptions if achieving the original DQOs is not realistic or practicable.

The last step (Step 7) in the DQO Process is described in detail in the next section of this document. Example applications of the full DQO Process are presented in Appendix "I."

[7] Under §261.38(c)(8)(iii)(A), a generator must demonstrate that "each constituent of concern is not present in the waste above the specification level at the 95% upper confidence limit around the mean."

Table 5. Summary of Outputs of the First Six Steps of the DQO Process

DQO Step	*Expected Outputs*
1. State the Problem	• List of members of the planning/scoping team and their role/expertise in the project. Identify individuals or organizations participating in the project (e.g. facility name) and discuss their roles, responsibilities, and organization. • A concise description of the problem. • Summary of available resources and relevant deadlines.
2. Identify the Decision	• A decision statement that links the principal study question to possible actions that will solve the problem or answer the question.
3. Identify Inputs to the Decision	• A list of informational inputs needed to resolve the decision statement, how the information will be used, sources of that information, and an indication of whether the information is available for will need to be obtained. • A list of environmental variables or characteristics that will be measured.
4. Define the Boundaries	• A detailed description of the spatial and temporal boundaries of the problem (i.e., define the population, each decision unit, and the sample support). • Options for stratifying the population under study. • Any practical constraints that may interfere with the study.
5. Develop a Decision Rule	• The parameter of interest that characterizes the population. • The Action Level or other method for testing the decision rule. • An “if ...then...” statement that defines the conditions that would cause the decision maker to choose among alternative actions.
6. Specify Limits on Decision Errors	• Potential variability and bias in the candidate sampling and measurement methods • The baseline condition (null hypothesis) • The boundaries of the gray region • The decision maker’s tolerable decision error rates based on a consideration of consequences of making an incorrect decision.