FOREWORD

The U.S. Environmental Protection Agency has developed this guidance as part of its Quality System, an Agency-wide program of quality assurance for environmental data. One component of this Quality System is the requirement that investigators use a systematic planning process as mandated in EPA Order 5360.1 CHG 1: *Policy and Program Requirements for the Mandatory Agency-wide Quality System* (EPA, 1998b). EPA strongly recommends the Data Quality Objectives (DQO) Process as the appropriate systematic planning process for decision making. The DQO Process is an important tool for project managers and planners to define the type, quality, and quantity of data needed to make defensible decisions.

Data Quality Objectives Process for Hazardous Waste Site Investigations (QA/G-4HW) is based on the principles and steps developed in *Guidance for the Data Quality Objectives Process (QA/G-4)* (EPA, 1994b) but is specific to hazardous waste site investigations. This guidance is also consistent with *Data Quality Objectives Process for Superfund: Interim Final Guidance* (EPA, 1993) and *Soil Screening Guidance: User's Guide* (EPA, 1996a). Although this document focuses on EPA applications, such as site assessments under the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) and the Resource Recovery and Conservation Act (RCRA), this guidance is applicable to programs at the state and local level.

This publication is one of the *U.S. Environmental Protection Agency Quality System Series* documents. These documents describe the EPA policies and procedures for planning, implementing, and assessing the effectiveness of the Quality System and provide suggestions and recommendations for using the various components of the Quality System.

- *Data Quality Objectives Decision Error Feasibility Trials (DEFT) Software (QA/G-4D)* (EPA, 1994c)
- *Guidance for Quality Assurance Project Plans (QA/G-5)* (EPA, 1998c)
- *Guidance for Data Quality Assessment: Practical Methods for Data Analysis (QA/G-9)* (EPA, 1996b)
- *Data Quality Evaluation Statistical Toolbox (DataQUEST) (QA/G-9D)* (EPA, 1997)

These and other related documents are available on the EPA's Quality Staff's Web site, www.epa.gov/quality. Questions regarding this or other available system series documents may be directed to:

Quality Staff (2811R)
U.S. Environmental Protection Agency
Washington, DC 20460
(202) 564-6830
FAX (202) 565-2441
E-mail: quality@epa.gov

United States
Environmental Pr tion
Agency

Office of Environmental
Information
Washington, DC 20460

EPA/600/R-00/007
January 2000

ER ID 098258

EPA

Data Quality Objectives Process for Hazardous Waste Site Investigations

EPA QA/G-4HW
Final

30842

TABLE OF CONTENTS

Page

FOREWORD . . . . . **i**

LIST OF FIGURES . . . . . **v**

LIST OF TABLES . . . . . **v**

LIST OF ACRONYMS . . . . . **vi**

CHAPTER 0. INTRODUCTION **1**
- 0.1 PURPOSE AND SCOPE OF THIS DOCUMENT 1
- 0.2 RATIONALE FOR THE DOCUMENT 1
- 0.3 INTENDED AUDIENCE 2
- 0.4 THE DQO PROCESS 2
- 0.5 THE DQO PROCESS APPLIED TO RCRA CORRECTIVE ACTION AND SUPERFUND 6
- 0.6 SPECIAL CONSIDERATIONS FOR DIFFERENT MEDIA 10
- 0.7 ORGANIZATION OF THIS DOCUMENT 12

CHAPTER 1. STEP 1: STATE THE PROBLEM **13**
- 1.1 BACKGROUND 13
- 1.2 ACTIVITIES 13
- 1.3 OUTPUTS 18

CHAPTER 2. STEP 2: IDENTIFY THE DECISION **19**
- 2.1 BACKGROUND 19
- 2.2 ACTIVITIES 19
- 2.3 OUTPUTS 21

CHAPTER 3. STEP 3: IDENTIFY THE INPUTS TO THE DECISION **23**
- 3.1 BACKGROUND 23
- 3.2 ACTIVITIES 24
- 3.3 OUTPUTS 25

CHAPTER 4. STEP 4: DEFINE THE BOUNDARIES OF THE STUDY **27**
- 4.1 BACKGROUND 27
- 4.2 ACTIVITIES 28
- 4.3 OUTPUTS 33

CHAPTER 5. STEP 5: DEVELOP A DECISION RULE 35
- 5.1 BACKGROUND 35

TABLE OF CONTENTS—CONTINUED

Page

5.2 ACTIVITIES 36
5.3 OUTPUTS 39

CHAPTER 6. STEP 6: SPECIFY TOLERABLE LIMITS ON DECISION ERRORS . . 41
6.1 BACKGROUND 41
6.2 ACTIVITIES 44
6.3 OUTPUTS 50

CHAPTER 7. STEP 7: OPTIMIZE THE DESIGN FOR OBTAINING DATA 53
7.1 BACKGROUND 53
7.2 ACTIVITIES 56
7.3 OUTPUTS 61

CHAPTER 8. BEYOND THE DQO PROCESS: QUALITY ASSURANCE PROJECT PLANS AND DATA QUALITY ASSESSMENT 63
8.1 OVERVIEW 63
8.2 THE PROJECT LIFE CYCLE 63

REFERENCES 69

APPENDICES

APPENDIX A. A COMPARISON OF DQO PROCESS DOCUMENTS A-1

APPENDIX B. GLOSSARY OF TERMS USED IN THIS DOCUMENT B-1

APPENDIX C. JUDGMENTAL SAMPLING DQO CASE STUDY:
ACCONADA STORAGE FACILITY C-1

APPENDIX D. PROBABILISTIC SAMPLING: DQO CASE STUDY:
BLUE MOUNTAIN SMELTER D-1

LIST OF FIGURES

		Page
Figure 1.	The Data Quality Objectives Process	2
Figure 2.	Comparison of Phases of Hazardous Waste Site Investigations Between the RCRA Corrective Action Program and Superfund	7
Figure 3.	Example of a Conceptual Site Model (CSM) Diagram	16
Figure 4.	Example of Defining Spatial Boundaries for a Soil Contamination Problem	29
Figure 5.	An Example of a Decision Performance Goal Diagram — Baseline Condition: Parameter Exceeds Action Level	45
Figure 6.	An Example of a Decision Performance Goal Diagram — Baseline Condition: Parameter Is Less Than Action Level	45
Figure 7.	Ideal Versus Realistic Power Curve	60
Figure 8.	An Example of a Performance Curve Overlaid on a Decision Performance Goal Diagram (Baseline Condition: Parameter Exceeds Action Level)	60
Figure 9.	The DQO Process Is the Initial Component of the Project Level of EPA's Quality System	63
Figure 10.	The Iterative Nature of the DQO Process	64
Figure 11.	Data Quality Assessment	68

LIST OF TABLES

		Page
Table 1.	Example Principal Study Questions	20
Table 2.	Example Alternative Actions	21
Table 3.	Example Decision Statements	22
Table 4.	Example Inputs for a Site Investigation Decision	26
Table 5.	Examples of Scales of Decision Making	32
Table 6.	Examples of a Decision Rule	39
Table 7.	Decision Error Limits Table Corresponding to Figure 5	46
Table 8.	Decision Error Limits Table Corresponding to Figure 6	46
Table 9.	Probabilistic Sampling Designs	55
Table 10.	Common Sample Size Formulas	58
Table 11.	QA Project Plan Elements	65

LIST OF ACRONYMS

ARARs	Applicable or Relevant and Appropriate Requirements
CA	Corrective Action
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
CERCLIS	Comprehensive Environmental Response, Compensation, and Liability Information System
CSM	Conceptual Site Model
DPGD	Decision Performance Goal Diagram
DQA	Data Quality Assessment
DQO or DQOs	Data Quality Objectives
EU	Exposure Unit
FS	Feasibility Study
FSP	Field Sampling Plan
HHEM	Human Health Evaluation Manual
NPL	National Priorities List
PCBs	Polychlorinated Biphenyls
PCE	Perchloroethylene
PRGs	Preliminary Remediation Goals
RA	Remedial Action
RAGS	Risk Assessment Guidance for Superfund
RCRA	Resource Conservation and Recovery Act
RD	Remedial Design
RFA	RCRA Facility Assessment
RFI	RCRA Facility Investigation
RI	Remedial Investigation
RU	Remediation Unit
SAFER	Streamlined Approach for Environmental Restoration
SAP	Sampling and Analysis Plan
SI	Site Inspection
SWMU	Solid Waste Management Unit

CHAPTER 0

INTRODUCTION

0.1 PURPOSE AND SCOPE OF THIS DOCUMENT

Data Quality Objectives Process for Hazardous Waste Site Investigations (QA/G-4HW) provides general, nonmandatory guidance on developing Data Quality Objectives (DQOs) for environmental data collection operations in support of hazardous waste site investigations. Application of the DQO Process will help site managers plan to collect data of the right type, quality, and quantity to support defensible site decisions.

This document focuses on planning for the collection of environmental measurement data in support of the more intensive investigations conducted under the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA or "Superfund") and the Resource Conservation and Recovery Act's (RCRA's) Corrective Action (CA) program, such as RCRA Facility Investigations (RFIs) and Superfund Remedial Investigations (RIs). Persons conducting hazardous waste site investigations in other, non-regulatory situations, such as real estate transfers and brownfields redevelopment, may also benefit from using this guidance.

Although this guidance primarily addresses environmental data collection during intensive investigations such as RFIs and RIs, other stages of data collection operations during hazardous waste site investigations (e.g., site assessment phases, remedial operations) can find value in using this guidance. However, investigators may need to adapt the DQO Process to their specific problem. For example, during early site assessment phases, where investigators generally examine existing site information and conduct site reconnaissance, planning teams can benefit from the qualitative DQO steps, but may have to allow for a more liberal interpretation of the quantitative steps.

0.2 RATIONALE FOR THE DOCUMENT

The DQO Process can be applied to environmental data collection operations under a variety of situations. To address the wide range of planning needs in the environmental community, the U.S. Environmental Protection Agency's (EPA's) Quality Staff has developed several generic documents about the DQO Process: *Guidance for the Data Quality Objectives Process (QA/G-4)* (EPA, 1994b) and its related document, *Data Quality Objectives Decision Error Feasibility Trials (DEFT) Software (QA/G-4D)* (EPA, 1994c). The general guidance on the DQO Process presents basic guidance on the DQO Process for environmental decision making under a range of general problem types. DEFT is interactive software that determines the approximate number of samples and associated costs that would be needed to satisfy a set of DQOs. This document is tailored to hazardous waste site investigations. Use of the DQO Process satisfies the requirement for systematic planning of EPA Order 5360.1 CHG 1, *Policy and Program Requirements for the Mandatory Agency-wide Quality System*, (EPA, 1998b).

0.3 INTENDED AUDIENCE

This document was developed for persons involved in the management, investigation, or oversight of hazardous waste sites. To maximize the effectiveness of the document, users should consult the specific guidance and requirements of the program under which their site is being administered.

Prior to initiating the planning of a data collection event, all members of the DQO planning team should review this document. By becoming familiar with the steps and concepts of the DQO Process, team members will be better able to participate and contribute to the successful planning of the investigation. To ensure that all stakeholders (such as private citizens) have an understanding of the DQO Process, this guidance should be made available in public dockets.

0.4 THE DQO PROCESS

The DQO Process is a seven-step iterative planning approach used to prepare plans for environmental data collection activities (see Figure 1). It provides a systematic approach for defining the criteria that a data collection design should satisfy, including: when, where, and how to collect samples or measurements; determination of tolerable decision error rates; and the number of samples or measurements that should be collected.

DQOs, outputs of the DQO Process, are qualitative and quantitative statements that are developed in the first six steps of the DQO Process. DQOs define the purpose of the data collection effort, clarify what the data should represent to satisfy this purpose, and specify the performance requirements for the quality of information to be obtained from the data. These outputs are then used in the seventh and final step of the DQO Process to develop a data collection design that meets all performance criteria and other design requirements and constraints.

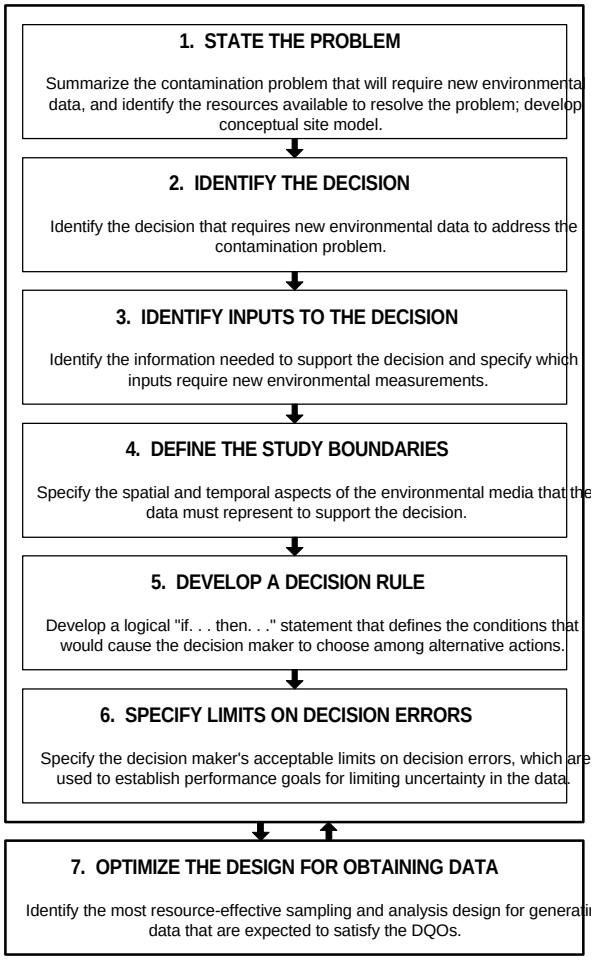


Figure 1. The Data Quality Objectives Process

In the context of a hazardous waste site investigation, a planning team may use the DQO Process at many stages of its involvement at

the site—from initial early assessments to site investigations and remedial operations. For example, a team may wish to determine whether or not a bioremediation technology has been effective in removing hazardous constituents from land-farmed sludge, and in particular, whether remediation should stop or continue for another year. There are risks involved in making the wrong decision in either case. If remediation halts before contaminant concentrations in the sludge have dropped below regulatory levels, then the land farm area may pose a hazard to human health and the environment. Conversely, if remediation continues when it is not needed, resources such as personnel and money will be spent needlessly. By using the DQO Process, the team members can clearly define what data and information about the bioremediation technology are needed; and they can develop a data collection design to help them obtain the right type, quantity, and quality of data they need to make a sound decision about whether the technology has been effective.

0.4.1 Planning and the EPA Quality System

EPA Order 5360.1 CHG 1: *Policy and Program Requirements for the Mandatory Agency-wide Quality System*, (EPA, 1998b), requires the use of a systematic planning process for all data collection and/or use by or for the Agency. The Order states that environmental data operations should be planned using a systematic planning process based on the scientific method. The planning process should have a common-sense, graded approach to ensure that the level of detail in planning is commensurate with the importance and intended use of the work and the available resources.

Elements of a systematic documented planning approach include:

- Identification and involvement of the project manager, sponsoring organizations, officials, project personnel, stakeholders, scientific experts, etc. (DQO Step 1);
- Description of the project goal, objectives, and issues to be addressed (DQO Steps 2 and 5);
- Identification of project schedule, resources, milestones, and any applicable regulatory and contractual requirements (DQO Step 2);
- Identification of the type of data needed and the ways in which the data will be used to support the project objectives and decisions (DQO Steps 3 and 4);
- Determination of the quantity of data needed and specification of performance criteria for measuring quality (DQO Step 6);
- Description of how, when, and where the data will be obtained (including existing data) and identification of any constraints on data collection (DQO Step 7).

While not mandatory, the DQO Process is the recommended planning approach for many EPA data collection activities, especially for the investigation of hazardous waste sites.

0.4.2 The DQO Process and EPA's Quality System at the Project Level

A project's life cycle comprises three phases: planning, implementation, and assessment. In the planning phase, site investigators specify the intended use of environmental data to be collected and plan the management and technical activities (e.g., sampling) needed to generate the data using the DQO Process. During the implementation phase, investigators put the plan developed in the first phase into action by constructing a QA Project Plan and collecting and analyzing samples (or measurements) in conjunction with QA and QC protocols. In the assessment phase, investigators evaluate the results of the sampling and analysis through Data Quality Assessment (DQA) to determine if the assumptions and performance requirements specified during planning were satisfied.

The DQO Process is flexible and iterative. Often, especially for more complicated sites, a larger planning team may be more efficient because a broader range of technical and stakeholder issues may arise. Regardless of the complexity of the site or the size of the planning team, it is common for the team to return to earlier steps to rethink the DQO outputs. These iterations through the earlier steps of the DQO Process can lead to a more focused design that can save resources in later field investigation activities.

In Superfund, the outputs of the DQO Process are most often used during the RI to develop the sampling design for the Field Sampling Plan (FSP) and to prepare the QA Project Plan. The FSP and QA Project Plan are often combined to create the Sampling and Analysis Plan (SAP). In Superfund RIs, the SAP helps investigators ensure that data collection activities are consistent with previous data collection activities at the site. The SAP also provides a system for planning and approving field activities and is the basis for estimating the cost of data collection activities.

In RCRA Corrective Actions, the DQO Process is used most often during the RFI. Investigators use the outputs of the DQO Process to prepare the QA Project Plan for the RFI. Investigators then incorporate the QA Project Plan and the sampling design developed by the DQO Process into the RFI Workplan. In RCRA Corrective Action, site owners (or permittees) will most often be conducting the RFI. Therefore, the RFI Workplan allows a permittee to present to the oversight agency the permittee's plans to characterize the nature and extent of the release or contamination. As the RFI Workplan should meet with the oversight agency's approval, permittees are encouraged to use the DQO Process to demonstrate the defensibility of their data collection plan.

0.4.3 Benefits of the DQO Process

One important benefit of the DQO Process is that it provides investigators with a reliable methodology for clarifying how decisions about the site will be supported by environmental data and for establishing site-specific performance criteria for these decisions. In general, the DQO Process also:

- improves the application and interpretation of sampling designs by using statistical and scientific principles for optimization;
- addresses the right questions early in the investigation by obtaining better knowledge of the waste constituents;
- achieves efficiency through generating the appropriate type and amount of data necessary to answer the question;
- helps investigators conserve resources by determining which data collection and analysis methods are most appropriate for the data quality needs of the study; and
- provides investigators with a stopping rule—a way for the planning team to determine when enough data of sufficient quality have been collected to make site decisions with the desired level of confidence.

0.4.4 Statistical Aspects of the DQO Process

The DQO Process has both qualitative and quantitative aspects. The qualitative parts promote logical, practical planning for environmental data collection operations and complement the more quantitative aspects. The quantitative parts use statistical methods to design the data collection plan that will most efficiently control the probability of making an incorrect decision.

In general, the statistical procedures used in the DQO Process provide:

- a scientific basis for making inferences about a site (or portion of a site) based on environmental data;
- a basis for defining decision performance criteria and assessing the achieved decision quality of the data collection design;
- a foundation for defining QA and QC procedures that are more closely linked to the intended use of the data;
- quantitative criteria for knowing when site investigators should stop data collection (i.e., when the problem has been adequately characterized);

- a solid foundation for planning subsequent data collection activities; and
- a scientific and statistical basis to support the investigators' ensuing decision.

Although the statistical aspects of the DQO Process are important, planning teams may not be able to apply statistics to every hazardous waste site investigation problem. For example, in the early stages of site assessment [e.g., RCRA Facility Assessments, Superfund Preliminary Assessments/Site Inspections (PAs/SIs)], statistical data collection designs may not be warranted by program guidelines or site-specific sampling objectives. In some cases, investigators may only need to use judgmental sampling or make authoritative measurements to confirm site characteristics.

The media being investigated also may determine whether or not the use of statistical methods will be limited. For example, in ground water studies, investigators may locate monitoring wells based on prior knowledge of likely contaminant flow pathways instead of a purely statistical sampling design. The planning team should examine different aspects of the data collection problem and discuss whether statistical methods are needed with respect to the decisions being made and extent of inference desired. A discussion of these types of problems is presented in Section 0.6 of this guidance.

0.4.5 Availability and Need for Statistical Assistance

Planning teams that need assistance on the more complex statistical aspects of the DQO Process should consult an environmental statistician. However, guidance on statistical and sampling procedures may be found in *Guidance for Data Quality Assessment: Practical Methods for Data Analysis (QA/G-9)* (EPA, 1996b). Statistical books of environmental sampling and analysis include: *Statistical Methods for Environmental Pollution Monitoring* by Richard O. Gilbert (1987); *Statistics for Environmental Engineers* by Paul M. Berthouex and Linfield C. Brown (1994); *Geostatistical Error Management* by Jeffery C. Myers (1997); and *Environmental Statistics and Data Analysis* by Wayne R. Ott (1995). In addition, the Quality Staff also has developed a PC-based software, *Data Quality Evaluation Statistical Toolbox (DataQUEST) (QA/G-9D)* (EPA, 1997). DataQUEST helps investigators assess the data once it has been collected.

0.5 THE DQO PROCESS APPLIED TO RCRA CORRECTIVE ACTION AND SUPERFUND

0.5.1 Application of the DQO Process

The DQO Process may be applied to any environmental data collection activity performed at RCRA CA facilities or Superfund sites. Readers will generally find the DQO Process steps and activities in this guidance are most applicable during the RFI or RI.

In general, there are five elements to Superfund and RCRA CA programs (see Figure 2): initial site assessment, site investigation, evaluation of remedial alternatives, remedy selection, and remedy implementation.[1] Although there are differences between the administration and regulatory setting of the site assessment, site investigation, evaluation of remedial alternatives, remedy selection, and remedy implementation programs, one of EPA's current initiatives is to develop consistency between the policies and procedures of Superfund and RCRA CA. For further information on the changes proposed to RCRA CA and the program's relationship to Superfund, readers should consult *Federal Register* Vol. 55, No. 145, July 27, 1990 and *Federal Register* Vol. 61, No. 85, May 1, 1996.

Initial Site Assessment. In most cleanup programs, the first phase is an initial site assessment. The purpose of this activity is to gather information on site conditions, releases, potential releases, and exposure pathways. Investigators use this information to determine whether a cleanup may be required or to identify areas of concern for further study. Information collected during this phase usually forms the basis for determining whether the next stage, site investigation, is warranted.

	RCRA Corrective Action Program	Superfund
Initial Site Assessment	RCRA Facility Assessment (RFA)	Preliminary Assessment/Site Inspection (PA/SI)
Site Investigation	RCRA Facility Investigation (RFI)	Remedial Investigation (RI)
Evaluation of Remedial Alternatives	Corrective Measures Study (CMS)	Feasibility Study (FS)
Remedy Selection	Permit Modification or Amended Order	Record of Decision (ROD)
Remedy Implementation	Corrective Measures Implementation (CMI)	Remedial Design/Remedial Action (RD/RA); Remedy Operation and Maintenance

Figure 2. Comparison of Phases of Hazardous Waste Site Investigations between the RCRA Corrective Action Program and Superfund

[1]In addition, interim actions or emergency-response actions (e.g., stabilization, removal of wastes, institutional controls, supply of drinking water) may occur at any time during the program administration of a site or facility. Interim actions are used to control or minimize ongoing risks to human health and the environment.

In the RCRA CA program, the initial site assessment is called the RCRA Facility Assessment. EPA or a state authority conducts the RFA to determine whether there is any threat to human health and the environment at a facility. During the RFA, investigators identify and evaluate solid waste management units (SWMUs) and other areas of concern for releases to all media. In addition, investigators determine the need for further investigation and interim measures. If the facility poses a threat to human health or the environment, investigators may require corrective action either by a corrective action order or through the facility's permit conditions. For further guidance on the RFA, readers should consult *RCRA Facility Assessment (RFA) Guidance* (EPA, 1986).

In the Superfund program, this phase is called the Preliminary Assessment/Site Inspection. EPA or a state authority conducts a PA on a site listed in the Comprehensive Environmental Response, Compensation, and Liability Information System. The PA is generally limited in scope and consists of collecting available information and conducting a site reconnaissance. The purpose of the PA is to determine whether the site may pose a threat to human health and the environment. If investigators determine through the PA that further investigation is needed, then an SI will be initiated. During the SI, investigators usually collect environmental measurements to determine what hazardous substances are present at the site and whether or not they are being released to the environment. One objective of the SI is to provide a basis for ranking the site's hazards for possible placement of the site on the National Priorities List (NPL). A second objective of the SI is to determine if the site poses any immediate health or environmental risks and requires emergency response. For further information on the PA/SI, readers should consult *Guidance for Performing Preliminary Assessments Under CERCLA* (EPA, 1991a) and *Guidance for Performing Site Inspections Under CERCLA* (EPA, 1992a).

Site Investigation. The purpose of this phase is to determine the nature and extent of contamination at a site, quantify risks posed to human health and the environment, and gather information to support the selection and implementation of appropriate remedies.

In the RCRA CA program, this phase is known as the RCRA Facility Investigation. The facility owner or permittee generally conducts the RFI with oversight from EPA or a state authority. Through the RFI, the facility owner characterizes the nature, extent, direction, rate, movement, and concentration of releases at the facility as well as the chemical and physical properties of the site that are likely to influence contamination migration and cleanup. For further information on the RFI, readers should consult *RCRA Facility Investigation (RFI) Guidance* (Volumes I-IV) (EPA, 1989b), *RCRA Corrective Action Plan* (EPA, 1994a), and *Soil Screening Guidance: User's Guide* (EPA, 1996a).

In Superfund, this phase is referred to as the Remedial Investigation. RIs are conducted at sites placed on the NPL. EPA, state authorities, or potentially responsible parties may conduct RIs. During the RI, investigators define the nature and extent of contamination at the site and conduct a baseline risk assessment. For further information, readers should consult *Guidance for Conducting Remedial Investigations and Feasibility Studies Under CERCLA* (EPA, 1988), *Risk*

Assessment Guidance for Superfund: Volume I—Human Health Evaluation Manual, Part B, Development of Risk-Based Preliminary Remediation Goals (RAGS HHEM, Part B) (EPA, 1991c), and *Soil Screening Guidance: User's Guide* (EPA, 1996a).

Evaluation of Remedial Alternatives. The purpose of this phase is to assess the advantages and disadvantages of different potential remedial alternatives for the site or facility. In general, this stage is concurrent with either the RFI or RI, and investigators use data collected during the RFI or RI to develop options for remedial alternatives. In the RCRA CA program, this stage is known as the Corrective Measures Study. For more information on the Corrective Measures Study, readers should consult *RCRA Corrective Action Plan* (EPA, 1994a) and *RCRA Corrective Action Inspection Guidance Manual* (EPA, 1995a). In Superfund, this stage is the Feasibility Study (FS). For more information on the FS, readers should consult *Guidance for Conducting Remedial Investigations and Feasibility Studies Under CERCLA* (EPA, 1988).

Remedy Selection. During this stage, EPA selects a remedy for the site or facility that should be protective of human health and the environment, and should maintain that protection over time. In the RCRA CA program, either a permit modification or an amended order is issued by EPA or a State to support the selection of the final remedy. In Superfund, EPA prepares a Record of Decision to support the selection of the final remedy and documents data, analyses, and policy considerations that contributed to the remedy's selection.

Remedy Implementation. Remedy implementation consists of several activities: remedy design, remedy construction, remedy operation and maintenance, and remedy completion. In the RCRA CA program, these activities are known as Corrective Measures Implementation. In Superfund, these activities are called Remedial Design/Remedial Action (RD/RA) and Operation and Maintenance. Documentation for the remedy implementation should include the investigators' plans and methods to determine whether the remedy is effective and when remedial goals have been achieved.

0.5.2 Using This Document to Help Plan Studies

Planning teams should be familiar with the guidance before beginning the DQO Process and should document each step of the planning process, including all inputs and outputs. However, in some studies, investigators may not be able to complete Steps 6 and 7 in the manner described in the guidance. In these situations, investigators should always apply the fundamental underlying principles of the steps, base their data collection plans on some explicit consideration of tolerable uncertainty in the data, and document the reasons why the steps were not completed.

0.5.3 Other Guidance and Requirements Applicable to Investigations

This guidance provides nonmandatory instructions for applying the DQO Process to data collection activities at sites and facilities under RCRA Corrective Action or Superfund. Although this document has attempted to incorporate the programs' most current policies and guidelines,

readers should determine their program's latest requirements before conducting an investigation. In Section 0.5.1, documents useful for different stages of hazardous waste site investigations have been listed. Readers should refer to those documents as a starting point. To determine what guidance is the most appropriate and current, readers may wish to consult the RCRA/Superfund Hotline at (800) 424-9346, or in the Washington, DC, Metropolitan Area at (703) 412-9810.

0.6 SPECIAL CONSIDERATIONS FOR DIFFERENT MEDIA

This section contains a brief discussion of the different types of media that may be addressed in hazardous waste site investigations and, in general, some of the various problems one may encounter when applying the DQO Process. Note that this discussion is not an exhaustive list of considerations but is intended to give a sample of the types of issues and challenges that may arise. In all cases, the planning team should have scientific advisors who are experts in the media and conditions of the study.

0.6.1 Surface Soil

The various hazardous waste programs define surface soils differently depending on the purpose of the investigation and the exposure pathways for surface soils. In general, surface soils are considered to be the top 1 inch (or 2 centimeters) of soil. (However, under certain conditions, some programs alternatively define surface soil as the top *6 inches* of soil. Readers should determine their program's requirements.) Development of DQOs for surface soil investigations is generally straightforward because of the relative ease in preparing a statistical sampling design in a medium that is more stable, static, and readily bounded than other media. In fact, readers will find that the majority of examples of the DQO Process are presented in the surface soil medium.

However, planning teams may encounter a few problems in the application of the DQO Process to surface soils. For example, site surface soils may be extremely heterogeneous (e.g., soils with a wide range of particle sizes from clays and silts to cobbles, and even wastes such as plastic scrap or fiberglass insulation). Because contamination adheres differently to the various components of the soil and debris, investigators will have to consider how to develop a sampling design that will collect measurements that are truly representative of the media and the contamination. In addition, a highly heterogeneous surface soil presents problems in the actual physical sampling of the media. Investigators should determine what methods are most appropriate for the physical characteristics of the site. For more information on surface soil sampling considerations, readers should consult *Soil Screening Guidance: User's Guide* (EPA, 1996a).

0.6.2 Subsurface Soil

Subsurface soils present a problem to investigators because the soils are difficult to characterize fully. By most definitions, subsurface soils represent the soil media from approximately 1 inch below the ground surface to the top of the water table. When using the

alternate definition of surface soil, the subsurface soil represents the soil media from 6 inches below the ground surface to the top of the water table. This zone can be a few inches or a few tens of feet in thickness. The characterization of subsurface soils is important because the soils may affect other media significantly. Contaminants from this zone can migrate to the surface or to ground water, where contaminants may pose a risk to human health. For a thick subsurface soil, sampling can be very expensive, requiring mobilization of drill crews and collection and analysis of deep soil cores. In addition, practical considerations such as concern about transferring contamination to lower soil zones can limit the number of samples taken in the subsurface soil. Because of these constraints and the natural variability of the subsurface, planning teams can be faced with a great deal of uncertainty in their subsurface soil data.

The science of and methods used in subsurface investigations are evolving continually. An elementary example of the application of the DQO Process to subsurface soils may be found in *Soil Screening Guidance: User's Guide* (EPA, 1996a), and a more complex discussion of soil sampling in general in Myers (1997).

0.6.3 Ground Water

Ground water is difficult to characterize because aquifers can be geographically and vertically extensive and complex. In addition, because ground water is usually flowing, investigators should be concerned with the temporal boundaries when defining a ground water population to characterize. Most planning teams encounter problems when trying to develop a statistical sampling design for ground water investigations. Investigators have developed some innovative approaches to this dilemma. For example, to determine whether a contaminant source has impacted ground water, investigators may use a statistical analysis of well measurements upgradient and downgradient from the source. For determining whether or not a ground water pump-and-treat technology is effective, investigators may use a statistical time series analysis of ground water data to assess whether contaminant concentrations are decreasing significantly. A statistical approach also may be used for locating wells along a point of compliance to ensure that a plume migrating past that point is detected with a specified level of confidence.

For further information on ground water monitoring, readers may wish to consult *Considerations in Ground-Water Remediation at Superfund Sites and RCRA Facilities* (EPA, 1991b), *Guidance Document on the Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities* (EPA, 1989), and *Methods for Evaluating the Attainment of Cleanup Standards, Volume 2: Ground Water* (EPA, 1992b).

0.6.4 Surface Water

In surface water investigations, the planning team's objective is generally to characterize the nature, extent, and rate of migration of contaminants to the medium. Like ground water, surface water can be difficult to characterize because of its three dimensions and its variation over time. However, surface water is easier to access for measurements than ground water.

Investigators can often monitor streams and lakes at key locations. Usually, surface water investigations will require the characterization of not only the water itself but also the bottom sediments and biota of the environment. The dynamics of sediment analysis with the problem of thin stratification can be complex. Depending on the hydrologic system, contaminants from the ground water may also affect surface water. For further information, readers should consult *RCRA Facility Investigation Guidance (RFI), Volume III, Air and Surface Water Releases* (EPA, 1989) and *Guidance for Conducting Remedial Investigations and Feasibility Studies Under CERCLA* (EPA, 1988).

0.6.5 Air

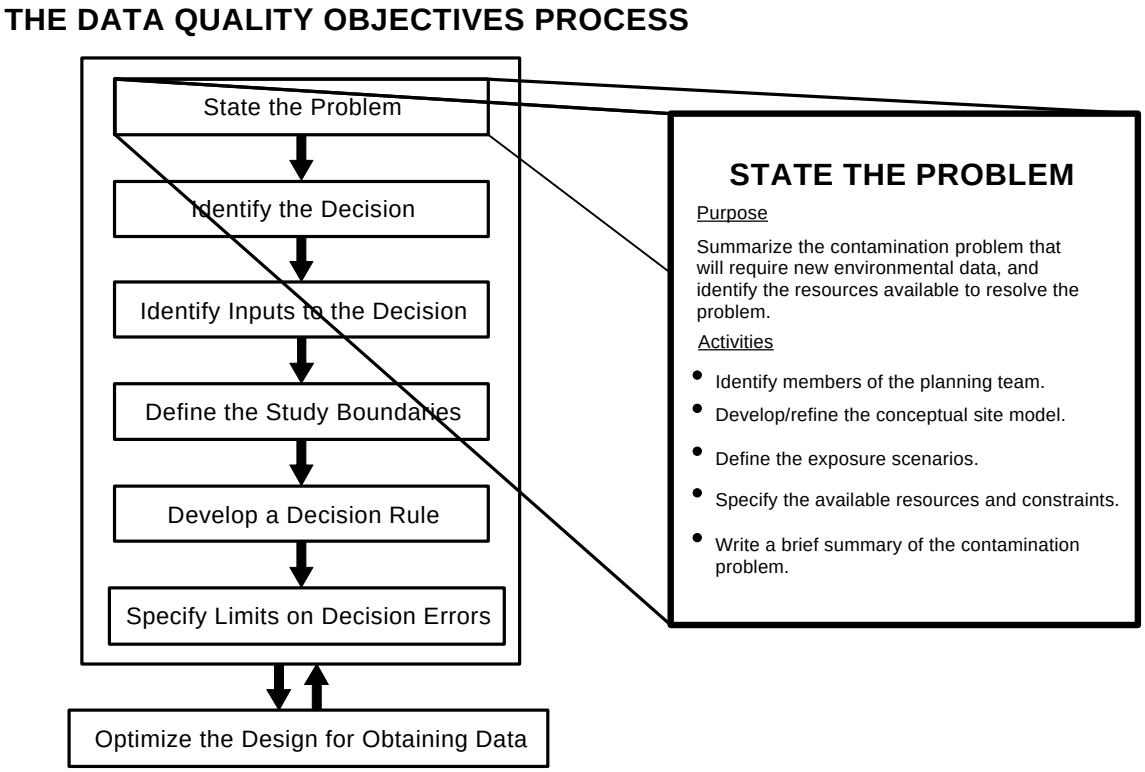
Air is difficult to characterize because investigators should consider how to collect data on a three-dimensional medium whose properties can change rapidly over time. Meteorological conditions such as wind speed and direction can greatly affect the concentrations of contaminants present in the air. In most cases, investigators will be concerned with contaminants such as volatile organics and airborne particulates, possibly being released to the environment from surface impoundments, landfills, or contaminated soils. Often, the planning team will need to determine whether air contaminants are present at the site or facility boundary. Generally, a monitoring network is set up along this boundary or models developed to predict exposure. Readers should consult *RCRA Facility Investigation Guidance (RFI), Volume III, Air and Surface Water Releases* (EPA, 1989) and *Guidance for Conducting Remedial Investigations and Feasibility Studies Under CERCLA* (EPA, 1988) for more information.

0.7 ORGANIZATION OF THIS DOCUMENT

Chapters 1 through 7 describe procedures for implementing the DQO Process at hazardous waste sites. Each chapter describes a step of the DQO Process, provides background material on the purpose of the step, and discusses the activities that produce the DQO outputs. Chapter 8 describes some of the more important activities following the completion of the DQO Process. This guidance is supported by several appendices. Appendix A compares three different documents that present versions of the DQO Process—*Guidance for the Data Quality Objectives Process EPA (QA/G-4)* (EPA, 1994b), the Department of Energy's "Streamlined Approach for Environmental Restoration (SAFER)" from its *Remedial Investigation/Feasibility Study (RI/FS) Process, Elements, and Technical Guidance* (DOE, 1993), and the American Society for Testing and Materials (ASTM) *Standard Practice for Generation of Environmental Data Related to Waste Management Activities: Development of Data Quality Objectives* (ASTM, 1996). Appendix B contains a glossary of terms used in this guidance, Appendix C is a DQO Case Study involving judgmental sampling schemes, and Appendix D is a DQO Case Study involving probabilistic sampling.

CHAPTER 1

STEP 1: STATE THE PROBLEM



1.1 BACKGROUND

The DQO Process may be applied to the investigation of contamination problems at hazardous waste sites during different phases—from initial site assessment activities to evaluations of remedial operations. By using the DQO Process, the site manager and the planning team can develop a framework for addressing specific contamination problems and determine sampling designs that are intended to collect the right type, quantity, and quality of data to support decision making.

This step encourages site managers to consider the broad context of the problem so that important issues are not overlooked. Step 1 activities include forming a description of the contamination problem, defining the planning team and determining organizational and management issues (e.g., determining members' roles, financial resources, and constraints).

1.2 ACTIVITIES

The three most important activities are to:

- describe the contamination problem that presents a potential threat or unacceptable risk to human health and the environment, and
- establish the DQO planning team,
- identify resources and organization/management issues needing resolution.

1.2.1 Identify Members of the Planning Team

The DQO planning team usually includes the site manager, regulatory authorities, and associated technical staff, together with stakeholders from the local community if appropriate.

The *site manager*[2] is typically the decision maker for the site and should actively participate in DQO development but may delegate responsibility for accomplishing planning tasks to the other members of the team. The decision maker also makes the final determination on the tolerable probability for the risk of decision errors in Step 6 of the DQO Process.

Regulatory Authorities are entities having policy inputs to the decision to be made. For example, State environmental organizations, EPA Regional staff, or local jurisdictions that require their viewpoints to be incorporated into the process to ensure a successful conclusion.

The *technical staff* should include representatives who are knowledgeable about technical issues that may arise over the course of several project phases. Depending on the nature of the contamination problem, the planning team of multidisciplinary experts may include QA specialists, samplers, chemists, modelers, technical project managers, human health and ecological risk assessors, toxicologists, biologists, ecologists, geologists, soil scientists, engineers, executive managers, data users or statisticians.

Stakeholders may consist of interested persons from the local community, such as nearby residents, local government authorities, and local businesses concerned with contamination problems and subsequent activities at the site.

DQO development does not always require a large planning team that includes every available area of expertise. For small sites with familiar contamination problems, the site manager may want to complete DQO development with a small team consisting of, for example, an environmental engineer, sampling expert, and laboratory manager. However, as the DQO Process is iterative, further experts can be added as the problem becomes more fully developed.

---

[2]In the Superfund program, the decision maker will typically be the site manager, also known as the Remedial Project Manager (RPM). If the RPM is not the decision maker, the person with this authority should be identified. In the RCRA Corrective Action program, the facility's oversight agency will need to determine a decision maker, because "site managers" in this case typically will be facility operators or permit holders who do not have the authority to make decisions such as acceptable risk levels for the site.

1.2.2 Develop/Refine the Conceptual Site Model

A conceptual site model (CSM) is a functional description of the contamination problem. The CSM should be initiated at the start of a project and carefully maintained and updated throughout the life of the site activities. The CSM is often accompanied by a CSM diagram (Figure 3), which illustrates the relationships among:

- locations of contaminant/waste sources or locations where contamination exists,
- types and expected concentrations of contaminants,
- potentially contaminated media and migration pathways, and
- potential human and ecological targets or receptors.

The planning team initially develops the CSM by collecting all available historical site data, including QA and QC documentation associated with previous environmental data collection activities. Presenting historical site data in this manner provides a foundation for identifying data gaps and focuses on where the problems of potentially unacceptable contamination may or may not exist.

Most hazardous waste programs have certain specific steps for developing CSMs, and investigators should consult their program's requirements. For Superfund, planning teams should consult *Guidance for Performing Site Inspections Under CERCLA* (EPA, 1992a), *Guidance for Conducting Remedial Investigation and Feasibility Studies Under CERCLA* (EPA, 1988), and *Soil Screening Guidance: User's Guide* (EPA, 1996a). For RCRA Corrective Action, planning teams should refer to *Soil Screening Guidance: User's Guide* (EPA, 1996a) which provides a checklist for developing an extensive, detailed CSM that was developed for use in soil screening but that investigators may find helpful in preparing CSMs for hazardous waste sites in general.

1.2.3 Define the Exposure Scenarios

At hazardous waste sites, the goal of investigation activities is usually to define site conditions that indicate or could lead to an unacceptable threat or exposure to human or ecological receptors. Whereas the CSM developed previously describes *potential* pathways, the preliminary exposure scenario describes the set of pathways that are consistent with *future uses* or activities at the site. For Superfund sites in particular, future uses and activities at the site may be different from the site's current or past uses and activities. For example, a former tannery site may be designated for future residential use. In this scenario, former activities that might lead to exposure, such as site workers coming into contact with hazardous sludge, may no longer apply; rather, the planning team may have to consider different activities under which exposure may occur, such as children coming into contact with contaminants through ingesting soil.

Investigators should combine information on potential human and ecological receptors around the site with likely contaminant migration pathways to develop preliminary exposure scenarios. The extent and methods for defining the scenarios may also depend on program-

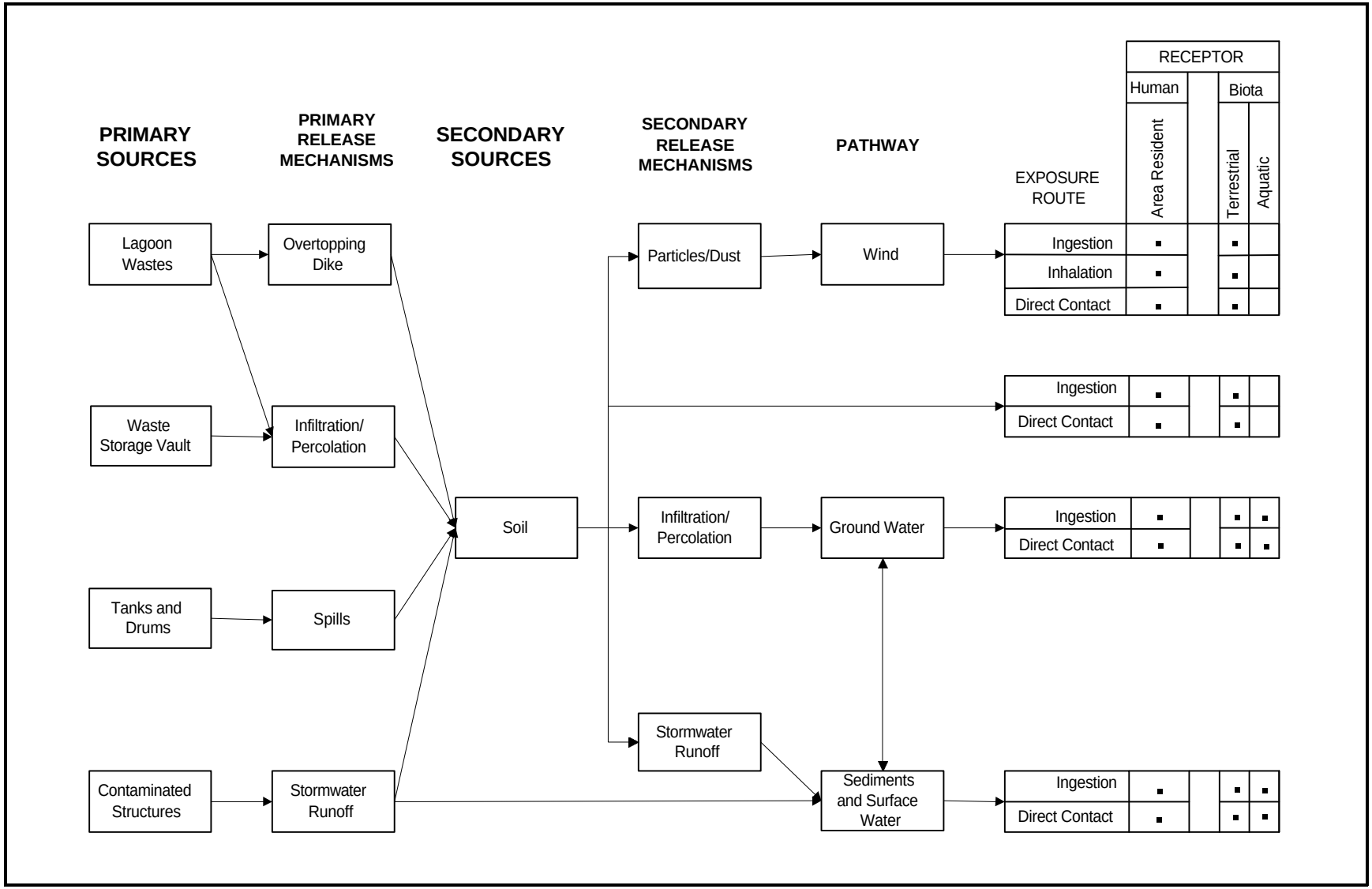


Figure 3. Example of a Conceptual Site Model (CSM) Diagram

specific requirements that the planning team should determine and consider when defining exposure scenarios.

For the early phases of investigation activities, it is necessary to establish which complete exposure pathways exist for each medium and land-use combination. In general, the planning team will:

- identify currently contaminated media to which individuals or sensitive ecosystems may be exposed;
- identify potential contaminants of concern based on historical site use, analytical data, and anecdotal information;
- define the current and future land use;
- determine the Applicable or Relevant and Appropriate Requirements (ARARs) for the site;
- for cases where multiple contaminants exist and ARARs are not available for all of the contaminants, develop risk-based contaminant-specific cleanup goals (for Superfund, these are called preliminary remediation goals or PRGs. Chemical-specific PRGs are concentrations based on ARARs or are concentrations based on risk assessment. Risk-based cleanup goals should also be developed for those contaminants for which meeting all ARARs is not considered protective); and
- identify available toxicity values for all the contaminants of concern and assemble these values along with the information obtained in the previous steps into exposure scenarios that should represent the highest exposure that could reasonably occur at the site.

More detailed information on accomplishing the above activities under Superfund can be found in *Risk Assessment Guidance for Superfund: Volume I—Human Health Evaluation Manual, Part B, Development of Risk-Based Preliminary Remediation Goals* (RAGS HHEM, Part B) (EPA, 1991c) and *Risk Assessment Guidance for Superfund: Volume II—Environmental Evaluation Manual* (RAGS EEM) (EPA, 1991d). Note that the models, equations, and assumptions presented in *Soil Screening Guidance: User's Guide* (EPA, 1996a) to address inhalation exposures supersede those described in RAGS HHEM, Part B, for residential soils. More information for completing these activities under RCRA Corrective Action may be found in *RCRA Facility Investigation Guidance, Volumes I-IV* (EPA, 1989) and in the *Federal Register*, Vol. 55, No. 145, July 27, 1990.

1.2.4 Specify the Available Resources and Constraints

The planning team should specify the approximate monetary budget for the data collection activity. This estimate should account for developing DQOs, constructing the QA Project Plan, and implementing the sampling (or taking measurements), chemical analysis activity, and data handling and interpretation phases. In addition, the planning team should specify available personnel, contractual vehicles (if available), and any other additional resources.

The planning team should also look at the "big picture" with respect to the total cost of investigation and cleanup activities at the site. For example, performing a more thorough and expensive data collection event at one stage of the investigation may provide the data needed to make decisions at later stages, thereby eliminating the need for an additional sampling round and possibly reducing the total cost of the investigation.

In this activity, the planning team also determines the time constraints (e.g., compliance with RCRA permits) for completing the required site evaluations. Other issues to consider may include political factors, such as public concern, and whether health and ecological risks are time critical.

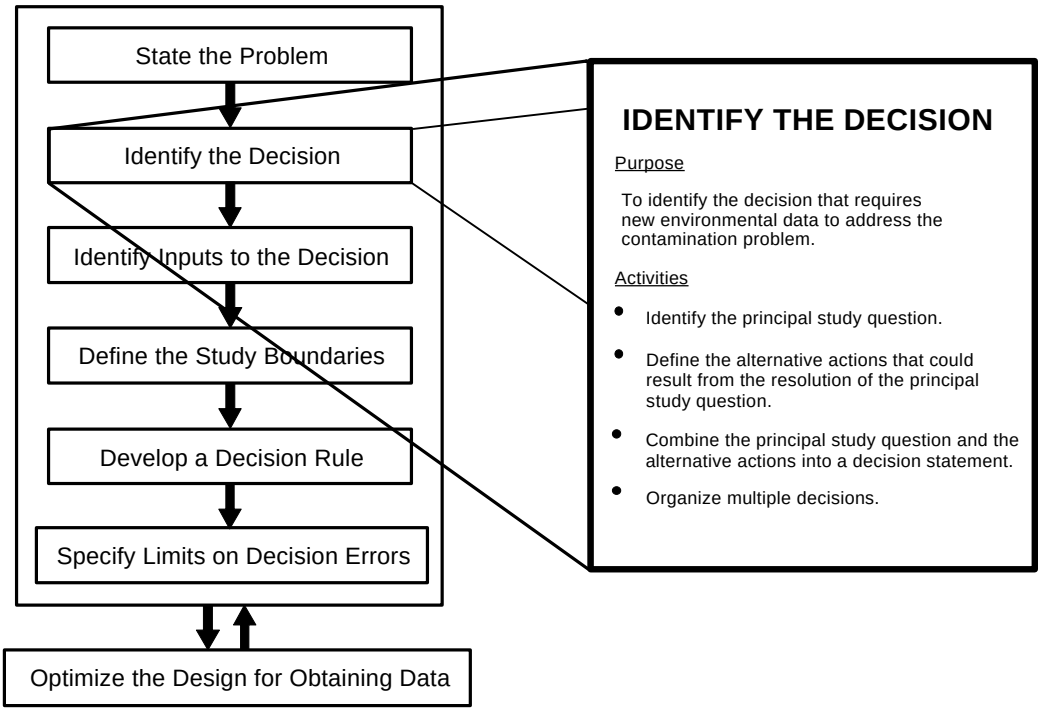
1.3 OUTPUTS

The main output of this step is a description of the contamination problem with its regulatory and programmatic context, the CSM and an estimate of the budget, schedule, and personnel necessary to implement the appropriate response for the site. The output should also identify the DQO planning team members and outline their most important responsibilities.

CHAPTER 2

STEP 2: IDENTIFY THE DECISION

THE DATA QUALITY OBJECTIVES PROCESS



2.1 BACKGROUND

The purpose of this step is to define the decision statement which combines the key question the study will attempt to resolve with the alternative actions that may be taken. In the DQO Process, the decision statement is abbreviated to simply "the decision."

2.2 ACTIVITIES

There are four activities in this step: identify the principal study question, define the alternative actions, combine the principal study question and alternative actions into a decision statement, and organize multiple decisions. Site managers usually address these activities in the order in which they appear in this chapter, but occasionally the team may wish to identify alternative actions before developing the principal study question. In some cases, the team will choose a decision statement specific to the requirements of the overall Agency or regulatory program.

2.2.1 Identify the Principal Study Question

The planning team reviews the problem stated in Step 1 and uses this information to identify the principal study question. The purpose of the principal study question is to allow investigators to narrow the scope of the search for information needed to address the problem. It is recommended that the initial iterations of the DQO Process concentrate on only one principal study question. Secondary study questions may be investigated in subsequent iterations. Some examples of principal study questions are provided in Table 1.

Table 1. Example Principal Study Questions

Stage	Principal Study Questions
Early Assessment Evaluations	Has a release of hazardous waste that poses a potential threat to human health or the environment occurred? Does the site contamination pose an unacceptable risk to human health or the environment?
Advanced Assessment Evaluations	Where do the contaminant concentrations exceed ARARs or exceed contaminant concentrations corresponding to the preliminary remediation goal for the site?
Assessment of Remedial Operations	Is the remedial technology performing at a level that will ensure remedial objectives are met?
Cleanup Attainment Evaluations	Has the final remediation level or removal action level been achieved?

2.2.2 Identify Alternative Actions that Could Result from the Resolution of the Principal Study Question

In this activity, the planning team identifies alternative actions that may be taken based on the outcome of the study and that correspond with the selected principal study question. The team will need to confirm that the actions associated with the decision will help resolve the contamination problem and determine if those actions are consistent with and satisfy the regulatory objectives. In addition, based on the statement of the problem and principal study question, investigators should verify that the actions help achieve the goal of protecting human health and the environment. Example alternative actions are provided in Table 2.

2.2.3 Combine the Principal Study Question and the Alternative Actions into a Decision Statement

In this activity, the team combines the alternative actions identified in the previous activity and the principal study question into a decision statement that presents a choice among alternative

Table 2. Example Alternative Actions

Stage	Alternative Actions
Early Assessment Evaluations	(i) Recommend that the site requires no further evaluation; or (ii) Recommend that the site warrants consideration of further assessment or a possible response action.
Advanced Assessment Evaluations	(i) Recommend that the site requires no further evaluation; or (ii) Recommend that the site warrants a possible response action.
Assessment of Remedial Operations	(i) Recommend that the current remedial technology continues operation; or (ii) Recommend that a new remedial technology or modifications to the current technology be considered.
Cleanup Attainment Evaluations	(i) Recommend that the site has achieved cleanup goals and proceed with delisting procedures; or (ii) Recommend that further response is appropriate for the site.

actions. The following standard form may be helpful in drafting decision statements: "Determine whether or not [environmental conditions/criteria from the principal study question] require (or support) [taking alternative actions]." Examples of decision statements are provided in Table 3.

2.2.4 Organize Multiple Decisions

If several separate decision statements should be defined to address the problem, the team should identify the relationships among the decisions and the sequence in which the decisions should be resolved. This activity may be regarded as placing the decision statements in an order of relative priority. The team may wish to document the decision resolution sequence and relationships in a diagram or flowchart.

2.3 OUTPUTS

The output of this step is a decision statement or set of statements that link the principal study question to possible or potential actions that will resolve the problem.

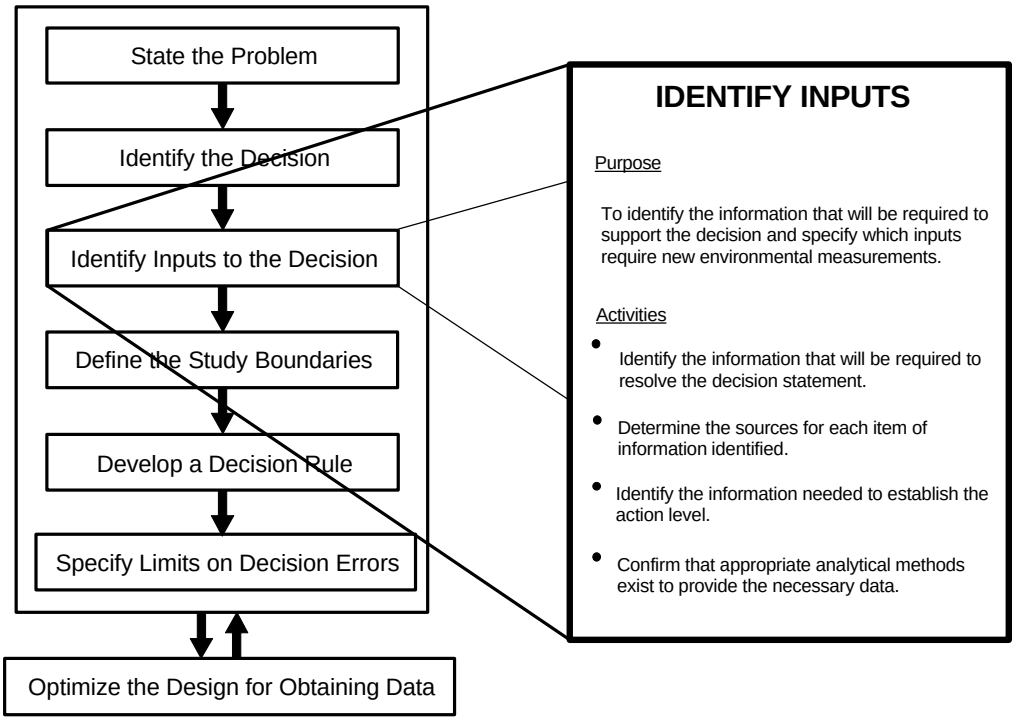
Table 3. Example Decision Statements

Stage	Decision Statements
Early Assessment Evaluations	Determine whether a release that poses a potential threat to human health and the environment has occurred and requires further consideration or a response action, or recommend that no further investigation is necessary. Determine whether site contamination poses an unacceptable risk to human health and the environment and requires further consideration or a response action, or recommend that no further investigation is necessary.
Advanced Assessment Evaluations	Determine where contaminant concentrations exceed ARARs or PRGs for the site and require further consideration or response action, and where no further investigation is necessary.
Assessment of Remedial Operations	Determine whether the remedial technology is attaining operational goals and should remain in operation, or whether a new technology or modifications to the current technology should be implemented
Cleanup Attainment Evaluations	Determine whether remedial objectives have been met such that no further action is required at the site and proceed with delisting procedures, or whether further response is appropriate for the site.

CHAPTER 3

STEP 3: IDENTIFY THE INPUTS TO THE DECISION

THE DATA QUALITY OBJECTIVES PROCESS



3.1 BACKGROUND

The purpose of this step is to identify the informational inputs needed to support the decision statement and to specify which inputs will require environmental measurements. This information is necessary so that the proper data may be collected to resolve the decision statement. To collect data that will be useful to resolve the decision statement, the planning team should identify what attributes are essential. The action level—such as an ARAR, a soil screening level (SSL), a PRG, or a RCRA Subpart S Action Level—is another important input that will be considered during this step. Once the planning team has determined what needs to be measured, the team will refine the specifications and criteria for the measurements in later steps of the DQO Process.

A conceptual understanding of the site (i.e., conceptual site model), as developed in Step 1, "State the Problem," which relates contaminant types and their sources to exposure pathways and receptors, is useful for identifying inputs. This conceptual site model and the decision statement defined in Step 2, "Identify the Decision," are previous outputs that are important to consider during this step.

3.2 ACTIVITIES

The following subsections describe activities that will help identify inputs to the decision.

3.2.1 Identify the Information That Will Be Required to Resolve the Decision Statement

The type of informational inputs necessary will depend on which approach is used to resolve the decision statement: sampling, modeling, or a combination of these approaches. For example, data on soil characteristics and hydrogeology are needed as inputs to model contaminant transport and dispersion through ground water in order to determine potential risks to receptors. The conceptual site model serves as a frame of reference for the data collection effort. Based on available data, the CSM summarizes how site-related contamination may pose a risk to human health and the environment. Some components of the conceptual site model may be estimated using mathematical equations and assumptions (i.e., modeling), and other components may be estimated by directly measuring some characteristic of the site (i.e., inference from a planned sampling study).

The analytical results of previous data collection activities should be summarized with respect to contaminants of interest; contaminant concentrations in each medium and the practical concentration ranges of concern; anticipated analytical methods; and analytical method performance characteristics (precision, bias, and method detection limits, etc.) to obtain a preliminary understanding of the problem.

A site visit or possibly a photographic site reconnaissance should be conducted (or the results from one recently completed should be obtained) to determine whether observations are consistent with the current understanding of the site. During this visit, the site should be searched for signs of contamination, such as discolored or odorous surface water, stressed vegetation, or discolored soil. Topographic maps should be used to mark locations and to estimate the extent of source areas or the presence of sensitive environs. The report should include information that will help assess the apparent stability of the site, such as leaking containment structures or weakening berms. Limited sampling should be conducted with portable equipment and additional anecdotal information gathered from local sources that may reveal disposal areas or practices that were previously unknown and may affect contaminant migration.

The planning team should list all information needed to resolve the decision statement. Diagraming techniques may help organize the inputs and show logical or temporal relationships.

3.2.2 Determine the Sources for Each Item of Information Identified

The planning team should identify existing sources for the informational inputs that will be required to resolve the decision statement. Sources may include historical records, regulations, directives, engineering standards, scientific literature, previous site investigations, professional

judgment, or new environmental measurements. Those inputs derived from new environmental measurements will be the main focus of subsequent DQO Process steps.

3.2.3 Identify the Information Needed to Establish the Action Level

The planning team will specify the basis for setting the action level.[3] The action level is the threshold value that provides the criterion for choosing between alternative actions. Action levels may be based on regulatory thresholds or standards, such as contaminant-specific ARARs or RCRA Subpart S Action Levels; they may be derived from site-specific risk considerations, such as RCRA media cleanup levels, PRGs, or soil screening levels; or they may be based on other criteria. If no existing source for action levels can be identified during this step, the site manager should decide how to develop a realistic concentration goal to serve as an action level for the field investigation design and evaluation. The goal of the current activity is merely to identify the regulatory or technical basis for setting the action level; the actual numerical value of the action level will be specified in Step 5, "Develop a Decision Rule." If the decision will be stated with respect to a background level, then instead of naming an action level, the team should identify where the background location will be chosen, and collect information on the characteristics of this background location. It is of great importance that the characteristics of the background location be compatible with those of the area under investigation.

3.2.4 Confirm that Appropriate Analytical Methods Exist to Provide the Necessary Data

The planning team should develop a list of potentially appropriate measurement methods for each item of necessary information. When data collection involves the chemical or biological sampling and analysis of environmental samples, it is preferable (if possible) to select a laboratory that is properly accredited to perform such analyses. Such laboratories are accredited through the National Environmental Laboratory Accreditation Program, which uses standards set by the National Environmental Laboratory Accreditation Conference. The main purpose here is to identify any situations where it may not be possible in practice to measure what is wanted. By identifying these situations early in the DQO Process, the planning team can consider other possible approaches, such as measuring surrogates, indicator variables, or adjustment of action levels to detection limits. Additional considerations about measurement detection limits are addressed in Step 5.

3.3 OUTPUTS

The outputs that will result from Step 3 activities include a list of informational inputs needed to resolve the decision statement and the sources of that information, including new environmental measurements. An example is given in Table 4.

---

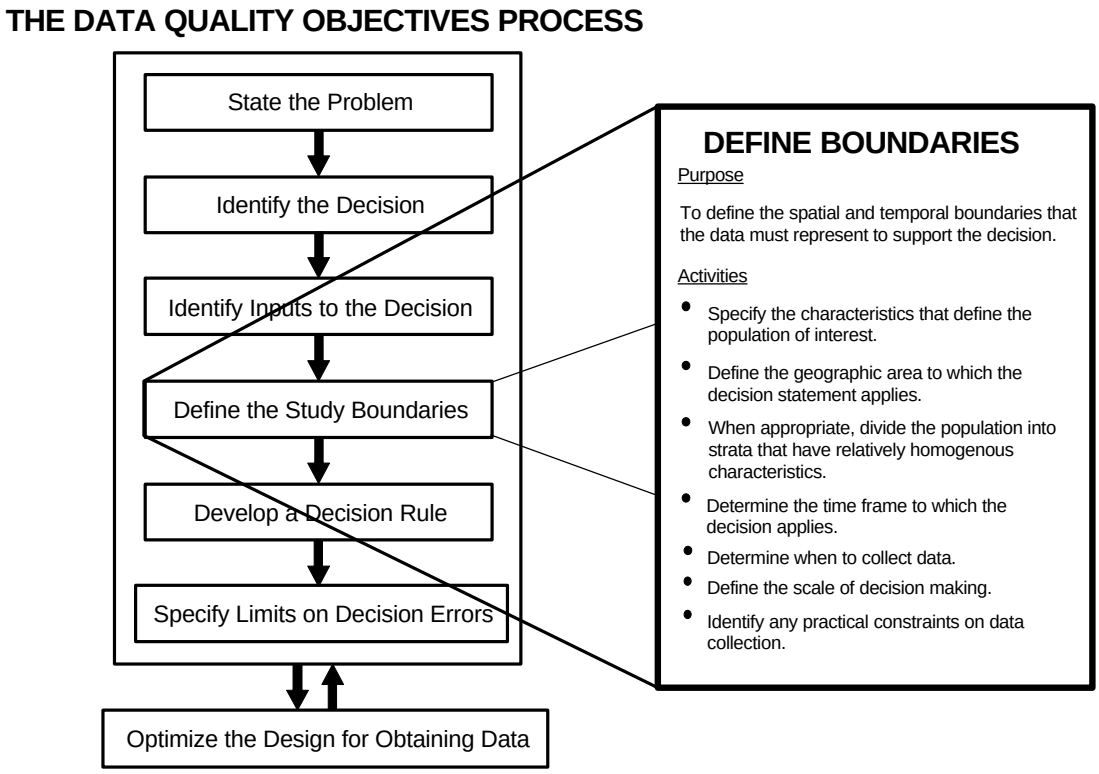
[3]In this document, the term "action level" refers to the value chosen in the DQO Process that provides the criterion for choosing between alternative actions. Readers will note that the RCRA Corrective Action program also uses the term "action level." To avoid confusion between the like terms, this document refers to action levels in the context of the RCRA CA program as "RCRA Subpart S Action Levels."

Table 4. Example Inputs for a Site Investigation Decision

Information Needed	Potential Source
Concentration values for arsenic, lead, and mercury in site soils	New environmental measurements (soil sampling and analysis)
Action level for each contaminant	Soil screening levels (SSLs) Preliminary remediation goal (PRG) calculations Record of Decision (ROD)

CHAPTER 4

STEP 4: DEFINE THE BOUNDARIES OF THE STUDY



4.1 BACKGROUND

The purpose of this step is to clarify the site characteristics that the environmental measurements are intended to represent. In this step, the planning team clearly defines the set of circumstances (i.e., spatial and temporal boundaries) that will be covered by the decision including:

- spatial conditions or boundaries of the site or release that define what should be studied and where samples should be taken, and

- temporal boundaries that describe what the time frame of the study data should be and when the samples should be taken.

Practical constraints that could interfere with sampling at the site also are identified in this step. The planning team should try to anticipate any obstacles that may interfere with the full implementation of the field sampling plan that will be developed from the DQOs and study design.

Applicable information from previous DQO steps that will be necessary to develop boundaries includes information from the conceptual site model developed in Step 1, "State the Problem," such as:

- site contaminants present or likely to be present and their potential sources;
- potential migration pathways, exposure routes, and receptors;
- the site's physical and chemical characteristics that affect contaminant distribution and enhance or decrease the likelihood of movement within and among media; and
- future use of the site.

This information is taken into account along with the decision statement or statements identified in Step 2, "Identify the Decision."

4.2 ACTIVITIES

The following subsections describe activities that provide details on specific portions of the boundaries step. Figure 4 illustrates schematically how boundaries may be defined for soil contamination problems. An accurate map of the site is critical.

4.2.1 Specify the Characteristics That Define the Population of Interest

The planning team should specify the characteristics that define the population of interest for the field investigation. The term "population" refers to the total collection or universe of objects, contaminated media, or people to be studied, from which samples will be drawn. It is important to clearly define the attributes that make up the population by stating them in a way that clarifies the focus of the study (for example, "2, 3, 7, 8-tetrachlorodibenzo-*p*-dioxin" (TCDD) is more specific than "dioxin"). In many cases, it is useful to state both the contaminant of concern and the matrix in which it is contained. For example, if a team is investigating lead contamination in soils at a site, the preferred specification of the population would be "lead contained in surface and subsurface soils." The possibility of intermedia transport also should be considered.

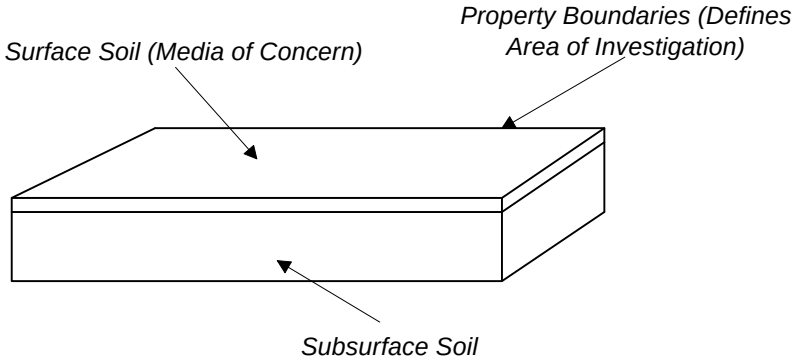
4.2.2 Define the Spatial Boundary of the Decision Statement

(1) **Define the geographic area and media to which the decision statement applies**. The geographic area is a region marked by some physical feature (e.g., volume, length, depth, width, political boundary) that limits the extent of the field investigation. Some examples of geographic areas are an operable unit of a Superfund site, the SWMU of a RCRA facility, the limits of a metropolitan city,

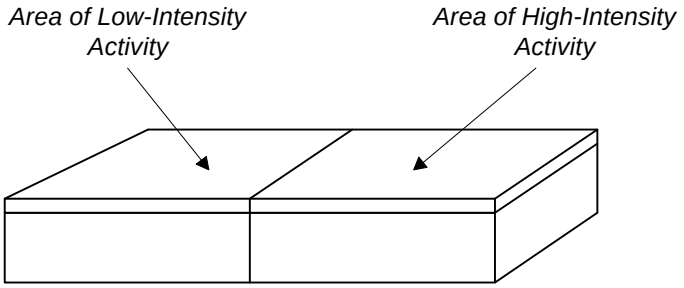
1. Define Population of Interest.

- Soil is the medium that is likely to be contaminated.
- Lead is the contaminant of concern.
- Intermedia transfer is not considered to be an important factor at this site.

2. Define Geographic Area of the Investigation and Media of Concern.



3. Stratify the Site.



4. Define the Temporal Boundaries of the Decision Statement.

- Time frame to which decisions apply (make a decision at the end of 4 years).
- When to collect data (sample every 6 months).

5. Define Scale of Decision Making.

- The scale of decision making for surface soil is based on risk exposure to residential families living on 1/2-acre lots.

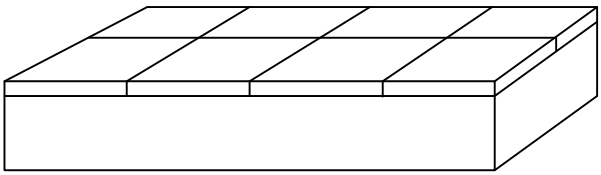


Figure 4. Example of Defining Spatial Boundaries for a Soil Contamination Problem

the property boundaries, and the natural habitat range of a particular animal species. The depth of the geographic area also should be included as this may bear on the selection of an action level.

(2) **When appropriate, divide the population into strata that have relatively homogeneous characteristics**. Using existing information, divide or stratify[4] the population or geographic area of the study into subsets or smaller areas that exhibit relatively homogeneous properties within each subset. Strata may be physically based, such as geological strata that affect contaminant distribution; or based on other factors, such as activity patterns that determine the likelihood of contamination. Stratification is desirable for studying subpopulations or for reducing the complexity of the problem by breaking it into more manageable pieces. It also can improve the efficiency of the sampling design. The site manager can then choose to make separate decisions about each stratum as well as the entire population.

4.2.3 Define the Temporal Boundaries of the Decision

(1) **Determine the time frame to which the study data apply**. It may not be possible to collect data over the full time period to which the decision will apply, particularly when long-term exposures are assumed in the future-use scenario. Therefore, the planning team needs to determine the most appropriate time frame that the data should represent (e.g., the study data will reflect the condition of the contaminant leaching into ground water over a period of 100 years) and determine a time frame for data collection that will best represent the full time period within the study constraints. Time frames should be defined for the overall population and for any subpopulations of interest. The planning team should note potential uncertainties due to mismatches between short time frames for sample collection versus long time periods to which the decision will apply.

(2) **Determine when to collect data**. Conditions may vary over the course of a study due to weather, seasonal variations, or other factors. For example, a study to measure exposure to volatile organic compounds from a contaminated site may give misleading information if the sampling is conducted in the colder winter months rather than in the warmer summer months. Therefore, the planning team should determine when conditions will be most favorable for collecting data and then select the time period that will reflect best the conditions of interest.

[4]Stratification is used to reduce the variability of contaminant concentrations and, therefore, to reduce the number of samples needed to meet the limits of decision error defined in Chapter 6.

4.2.4 Define the Scale of Decision Making

The scale of decision making is the smallest area or volume of the media, or the shortest time frame associated with the contamination problem of the site for which the planning team wishes to control decision errors. The goal of this activity is to define subsets of media about which the planning team will be able to make independent decisions that satisfy the decision error constraints specified in Step 6. The scale may range from the entire geographic boundaries of the site to the smallest area that can be remediated with a given technology. The scale of decision making is sometimes called a decision unit. The scale of decision making may be based on:

(1) **Risk**. The scale of decision making based on risk is determined by the relative exposure that an area presents to the receptor (i.e., the size of the decision unit is determined by the exposure scenario). The scale of decision making based on risk is referred to as an exposure unit (EU). An example of an EU is the ½ -acre residential lot used for the soil ingestion exposure route in *Soil Screening Guidance: User's Guide* (EPA, 1996a). Alternatively, the scale of decision making for the inhalation or migration to ground water exposure pathway is the entire contaminant source.

(2) **Permits/regulatory conditions.** A regulatory scale for decision making may be applied in RCRA Corrective Actions. The planning team may be required to make decisions for defined areas such as SWMUs.

(3) **Technological considerations**. A technological scale for decision making may be defined as the most efficient area or volume of the medium that can be removed or remediated with the selected technology. These areas or volumes are called remediation units (RUs). An example of an RU is the area of soil that can be removed by one pass of a bulldozer or the activities of a stationary backhoe.

(4) **Financial**. The financial scale is based on the actual cost to remediate that area of contaminated land. An extremely large EU, for example, may not be acceptable owing to the high cost of cleaning such a large area.

(5) **Other considerations**. Here, the scale of decision making is based on practical factors or on a combination of risk and technological factors that dictate a specific size. Examples are "hot spots," whose size may be based on historical site use and an acute exposure scenario. Examples of scales of decision making are included in Table 5.

A temporal scale of decision making might be necessary for studies where contamination varies significantly over time. For example, at a site with contaminated ground water, investigators may be concerned that quarterly sampling of perimeter monitoring wells might

Table 5. Examples of Scales of Decision Making

Scenario	Scale Chosen
Risk-Based: A lead smelter in Montana has contaminated approximately 35 acres with lead tailings and ash. The smelter site is surrounded by residential homes, and it is likely that the site could be used as residential lots in the future. The primary contaminant of concern on the site is lead in the soil, the exposure pathway is ingestion, and the primary target receptor is small children. One of the primary activities of children that exposes them to soil is playing in their backyards in play areas that are devoid of vegetation.	The planning team and the risk assessor want to control uncertainty in the sampling data related to the area or volume where children get the majority of their exposure. Therefore, the scoping team sets the scale of decision making to a 14' x 14' area, which is the average size of a backyard play area.
Technology-Based: A Midwestern coke plant has discharged process waste water into lagoons on its property, resulting in the contamination of sediments with organic chemicals. The lagoons are surrounded by a wetland area that is the primary concern as a receptor for the contamination, but there are no human receptors nearby. The cleanup of the lagoons will involve more than one type of remediation practice and is most likely to involve bioremediation and incineration to reduce the influence of the organic chemicals.	The planning team at this site chooses to evaluate each lagoon separately based on the assumption that each lagoon has homogeneous contamination that could be remediated by a single, but possibly separate, remediation process. Therefore, each lagoon is considered to be a distinct RU.
Other: The soil at an abandoned transformer production and reclamation facility has been contaminated with PCBs (polychlorinated biphenyls). The expected future use of the site is light industrial, and the major route of exposure is through soil ingestion. The site manager is most concerned with exposure to trespassing children who play on the site.	The planning team does not believe that there is a strong correlation between the size of a soil area and the relative amount of exposure that the children will receive. However, from the anticipated site activities of the children, they can select a size area (scale) that will be protective under the reasonable maximum exposure if that area had an average concentration of PCBs below the sampling and analysis action level. For this site, ½-acre is chosen as the scale of decision making. While this decision has to be based on some assumptions of risk and consideration of the receptor's activities, the planning team finally must estimate the size area that will protect the children rather than relying on a direct correlation between soil area and risk.

inadvertently allow rapidly migrating contamination to go undetected for too long and possibly endanger human health or the environment. Therefore, the investigators may choose a shorter period, such as a month, between sampling events.

4.2.5 Identify Any Practical Constraints on Data Collection

The team will identify any constraints or obstacles that could potentially interfere with the full implementation of the field sampling plan, such as weather conditions when sampling is not possible; the inability to gain access to sampling locations; or the unavailability of personnel, time, or equipment. For example, it may not be possible to take surface soil samples beyond one property boundary of a site because permission is not granted by the owner of the adjacent property.

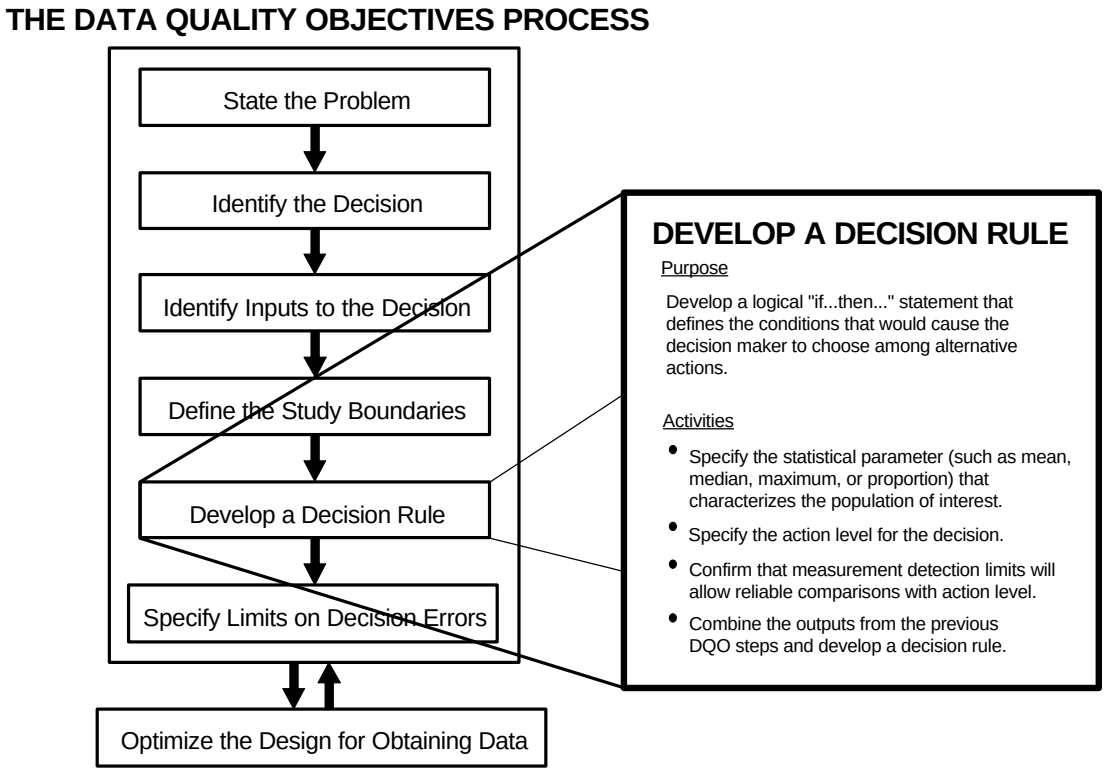
4.3 OUTPUTS

The outputs of this step are:

- a detailed description of the characteristics that define the population of interest;
- a detailed description and illustration of the geographic limits of each environmental medium (e.g., soil, water, air) within which the field investigation will be carried out;
- the time period in which samples will be taken and to which decisions will apply;
- the most appropriate scale of decision making for each medium of concern; and
- a description of practical constraints that may impede sampling.

CHAPTER 5

STEP 5: DEVELOP A DECISION RULE



5.1 BACKGROUND

In this step the planning team continues to build on the previous components of the decision-making framework established in earlier steps of the DQO Process. Specifically, the planning team:

- specifies the statistical parameter[5] that characterizes the population of interest;
- specifies the action level for the decision;
- confirms that the action level is above measurement detection limits so that reliable comparisons can be made; and

[5]The term "statistical parameter" refers to the key characteristics of the population of interest. By definition, it is unknown and can only be estimated by measuring a similar characteristic from a sample. For hazardous waste site investigations, the statistical parameters could be the overall mean level of contamination at the site, or upper 1 percent of contaminants (99th percentile) present on the site. It is standard practice to refer to population parameters using Greek letters, and their counterparts (sample statistics) by ordinary (Latin) letters.

- combines the statistical parameter, the scale of decision making, and the action level into an unambiguous decision rule that addresses the contamination problem.

The decision rule actually states what regulatory response action would be appropriate depending on whether the statistical parameter is greater or less than the action level. In practice, environmental data will be used to estimate the parameter but will almost surely differ from the true parameter value. Natural variability in data combined with the need to take a relatively small sample has created this unknown difference.

It is important to keep in mind that the decision rule in Step 5 is a "theoretical" decision rule that is stated in terms of what the decision maker ideally would like to know in order to choose the correct course of action. This activity is performed this way so that the DQOs are specified as generic performance requirements that allow flexibility in the statistical sampling design. In Step 5, the planning team members focus on what they would want to do if they could know with absolute certainty. One of the consequences of specifying the decision rule in these theoretical terms is that one need not address statements about the uncertainty of the parameter as part of the decision rule itself. The uncertainty that will apply to estimates of the parameter is addressed in Steps 6 and 7 of the DQO Process.

The decision rule combines the outputs from earlier steps, including the decision statement from Step 2, "Identify the Decision," the variables to be measured from Step 3, "Identify the Inputs to the Decision," and the scale of decision making from Step 4, "Define the Boundaries of the Study."

5.2 ACTIVITIES

5.2.1 Specify the Statistical Parameter That Characterizes the Population of Interest

The statistical parameter of interest is a descriptive measure (such as a mean, difference between two means, median, proportion, or maximum) that specifies the characteristic or attribute that the decision maker would like to know about the statistical population. In some cases, the study outcome or regulatory objectives state or imply a particular statistical parameter of interest; in other cases, it should be decided by the planning team.

The best guideline to follow when selecting a parameter of interest is to ask the question "What would I, the site manager, like to know?" If the answer is an average, then a mean or median might be selected. If the site manager would like to ensure that values in the population of interest fall below some concentration, a proportion or percentile should be used. If the site manager is interested in hot spots, then the maximum concentration or a certain diameter of hot spot might be a reasonable choice. If the site manager is interested in comparing the average between two populations (i.e., the site vs. background), then the parameter of interest is the difference between the mean of the site and the mean of the background. Choosing more complex parameters of interest (e.g., the third-highest maximum value) may lead to complex and

resource-intensive sampling designs in Step 7, "Optimize the Design for Obtaining Data," and should be avoided during the initial phases of DQO development.

The mean, a measure of central tendency in a population, is useful when the action level is based on long-term average health effects (e.g., chronic conditions, carcinogenicity) and when the population is fairly homogenous and has a relatively small variability (variance). Estimating the mean generally requires fewer samples than other parameters; however, the sample mean is not as good an estimate when the distribution underlying the population is highly skewed or when the population contains a large proportion of values that are less than the measurement method detection limit.

The population median is an alternative representation of the center of the population and is defined as that value where 50 percent of the values of the population are smaller than the median and 50 percent of the values are larger. Unlike the sample mean, the sample median is a good estimator of the center of a population that is highly skewed and can be used even if the population contains a large proportion of values that are less than the measurement detection level. However, because statistical tests concerning the median rely on fewer assumptions than do hypothesis tests concerning the mean, estimating the population median for use in statistical tests usually requires large sample sizes.

A proportion represents the number of objects in a population having (or not having) some characteristic divided by the total number of objects in the population. This characteristic may be qualitative, such as leaking drums versus nonleaking drums; or quantitative, such as drums with concentration levels of a contaminant greater than some fixed level. A proportion is useful if the population consists of discrete objects such as drums or finite units.

A percentile represents conditions where *x* percent of the distribution is less than or equal to the percentile value. For example, if the 95th percentile of a site is equal to 40 ppm, then 95 percent of the concentration levels at the site are less than or equal to 40 ppm. Statistical tests concerning percentiles are equivalent to those concerning proportions. Common population parameters at hazardous waste sites are upper percentiles (upper proportions) because they are conservative and protect against extreme health effects. A percentile provides controls for extreme values and is useful when the population contains a large number of values less than the analytical method detection limit. However, estimating upper percentiles for use in a statistical test usually requires large sample sizes.

5.2.2 Specify the Action Level for the Decision

The action level is a contaminant concentration or numerical value derived from ARARs or risk-based methodologies, such as the PRG development process, which, when applied to site-specific conditions, results in the establishment of a numerical criterion for deciding whether the contamination levels are unacceptable.

If the decision maker believes that the final remediation level could be one of two different levels, then the more stringent one should be chosen for the action level. A more stringent action level may require selection of more precise analytical methods—with appropriate detection limits— than would satisfy the less stringent action level; or it may require replicate analysis. In Superfund investigations, the planning team may need to develop PRGs and should refer to the *Risk Assessment Guidance for Superfund, Volume I—Human Health Evaluation Manual, Part B, Development of Risk-Based Preliminary Remediation Goals* (RAGS HHEM, Part B) (EPA, 1991c). Investigators should note that the models, equations, and assumptions used to develop risk-based action levels in *Soil Screening Guidance: User's Guide* (EPA, 1996a) supersede those described in RAGS HHEM, Part B, (EPA, 1991c) for residential soils. For RCRA Corrective Action, the planning team may need to develop media cleanup levels as discussed in the *Federal Register*, Vol. 55, No. 145, July 27, 1990, "Corrective Action for Solid Waste Management Units at Hazardous Waste Management Facilities; Proposed Rule."

There are several types of ARARs that remedial or removal actions may have to comply with. These include chemical-specific requirements that establish an acceptable residual amount or concentration of a contaminant, engineering design performance, action-specific requirements, or location-specific requirements. There are also nonpromulgated advisories or guidance documents that are not legally binding, referred to as "to-be-considered materials." In many instances, to-be-considered materials are part of risk assessment and are used to determine the level of cleanup necessary for health and environmental protection.

5.2.3 Confirm That the Action Level Exceeds Measurement Detection Limits

The planning team should examine the potential measurement methods identified in Step 3 and determine the detection limits for those methods. This performance information is used in this activity to confirm the feasibility of using that method to compare site concentrations to the action level. For example, if the detection limit exceeds the action level, then either a better method should be specified or a different approach should be used, such as measuring surrogates or indicators. This method performance information also will be used in Step 7, "Optimize the Design for Obtaining Data."

There are many different definitions of detection limits. The planning team should use the definition that is of most use for the decision rule at hand. For example, a decision rule that merely requires confirmation of the existence of a contaminant would require a detection limit that assumes a high probability of positive identification and presence in the matrix (and reasonably low probability of false confirmation). On the other hand, a decision rule that requires comparison of a mean contaminant concentration to a threshold action level value would require the detection limit to be defined in terms of the reliability of quantitation [such as a limit of quantitation or practical quantitation limit (PQL)].

5.2.4 Combine the Outputs from the Previous DQO Steps and Develop a Decision Rule

The planning team combines the decision statement, parameter of interest, scale of decision making, and action level into an "if. . .then. . ." statement that describes the conditions that would lead to a specific regulatory response action.

5.3 OUTPUTS

The output of this step is the "if...then..." decision rule. Examples of such a decision rule are shown in Table 6.

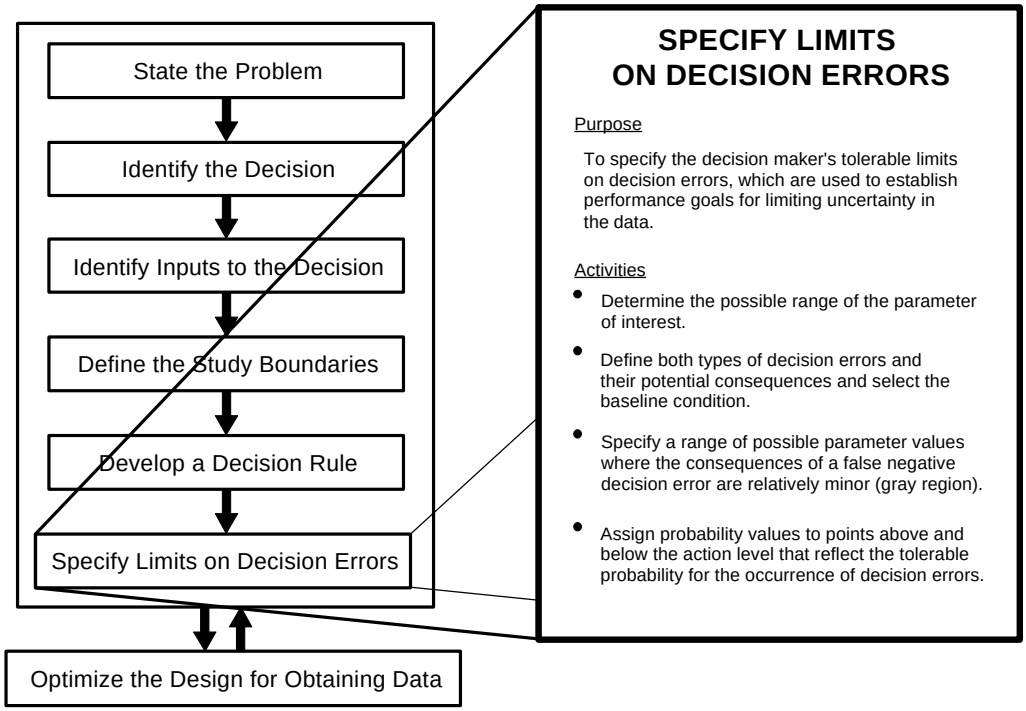
Table 6. Examples of a Decision Rule

If the mean perchloroethylene (PCE) concentration of each downgradient well is greater than the PCE concentration in an upgradient well, then further assessment and response are required; otherwise, no further evaluation is necessary.
If the mean level of arsenic is less than or equal to 1.0ppb, then the soil will be left in situ, otherwise the soil shall be removed to an approved site.

CHAPTER 6

STEP 6: SPECIFY TOLERABLE LIMITS ON DECISION ERRORS

THE DATA QUALITY OBJECTIVES PROCESS



6.1 BACKGROUND

The purpose of this step is to specify quantitative performance criteria for the decision rule expressed as probability limits on potential errors in decision making. The probability limits on decision errors specify the level of confidence the site manager desires in conclusions drawn from site data. These decision performance criteria will be used in Step 7, "Optimize the Design for Obtaining Data," to generate a resource-effective field investigation sampling design.

Setting tolerable limits on decision errors is neither obvious or easy. It requires the planning team to weigh the relative effects of threat to human health and the environment, expenditure of resources, and consequences of an incorrect decision, as well as the less tangible effects of credibility, sociopolitical cost, and feasibility of outcome. In the initial phases of the DQO development, these probabilities need only be approximated to explore options in sampling design and resource allocation. The effects of altering these probabilities on sampling plans and resources may be explored using the software, *Data Quality Objectives Decision Error Feasibility Trials (DEFT) Software* (EPA, 1994c).

6.1.1 Sources of Error in Hazardous Waste Site Investigations

A decision error occurs when the data mislead the site manager into choosing the wrong response action, in the sense that a different response action would have been chosen if the site manager had been able to access "perfect data" or absolute truth.

The possibility of a decision error exists because the parameter of interest is estimated using data that are never perfect but are subject to different variabilities at different stages of development, from field collection to sample analysis. The combination of all these errors is called "total study error," and for sampling at hazardous waste sites, this can be broken down into two main components:

(1) **Sampling design error**. This error (variability) is influenced by the sample collection design, the number of samples, and the actual variability of the population over space and time. It is impractical to sample every unit of the media, and limited sampling may miss some features of the natural variation of the contaminant concentration levels. Sampling design error occurs when the data collection design does not capture the complete variability within the media to the extent appropriate for the decision of interest.

(2) **Measurement error**. This error (variability) is influenced by imperfections in the measurement and analysis system. Random and systematic measurement errors are introduced in the measurement process during physical sample collection, sample handling, sample preparation, sample analysis, and data reduction.

In some cases, total study error may lead to a decision error. Therefore, it is essential to reduce total study error to a minimum by choice of sample design and measurement system in order to reduce the possibility of making a decision error.

6.1.2 Decision Making

The possibility of making a decision error, although small, is undesirable due to the adverse consequences arising from that incorrect decision. It can be controlled through the use of a formal statistical decision procedure, known as hypothesis testing. When hypothesis testing is applied to site assessment decisions, the data are used to choose between a presumed baseline condition of the environment and an alternative condition. The test can then be used to show either that the baseline condition is false (and therefore the alternative condition is true) or that there is insufficient evidence to indicate that the baseline condition is false (and therefore the site manager decides by default that the baseline condition is true). The burden of proof is placed on rejecting the baseline condition, because the test-of-hypothesis structure maintains the baseline condition as being true until overwhelming evidence is presented to indicate that the baseline condition is not true. For example, the site manager may presume that a site is contaminated (the

baseline condition) in the absence of strong evidence (data) that indicates the site is clean (the alternative condition).

A decision error occurs when the limited amount of data collected leads the site manager to decide that the baseline condition is false when it is true, or to decide that the baseline condition is true when it is really false. These two types of decision errors are classified as a false rejection error and a false acceptance error, respectively. In some circumstances, a false rejection error is known as a false positive error, and a false acceptance error as a false negative error. In statistical language, the baseline condition is called the null hypothesis (H_0) and the alternative condition is called the alternative hypothesis (H_a). A false rejection decision error occurs when the decision maker rejects the null hypothesis when it is really true; a false acceptance decision error occurs when the decision maker fails to reject the null hypothesis when it is really false.

Consider an example where the site manager strongly believes that the overall average level of contaminant of concern exceeds the action level (i.e., the baseline condition or null hypothesis states that COC concentrations exceed the action level). If the sampling data, by chance, contained an abnormally large proportion of low values, the site manager would erroneously conclude that the COC concentrations do not exceed the action level when in reality the true average did exceed the action level; the site manager would then be making a false rejection decision error.

Statisticians often refer to the false rejection decision error as a Type I error and the measure of the size of this error as alpha (α), the level of significance. Statisticians often refer to a false acceptance decision error as a Type II error; the measure of the size of this error is called beta (β), also known as the complement of the power of a test. Both alpha and beta are expressed numerically as probabilities.

6.1.3 Controlling Decision Errors

Although the possibility of decision errors can never be totally eliminated, it can be minimized and controlled. To control the possibility of decision errors, the planning team focuses on the largest components of total study error. If the sampling design error is believed to be relatively large, the chance of decision error may be controlled by collecting a larger number of samples or developing a better sampling design. If the analytical component of the measurement error is believed to be relatively large, it may be controlled by analyzing multiple individual samples, or by using more precise and accurate analytical methods.

In some cases, placing a stringent (i.e., very small) limit on the possibility of both types of decision errors is unnecessary for making a defensible decision. If the consequences of one decision error are relatively minor, it may be possible to make a defensible decision based on relatively imprecise data or on a small amount of data (e.g., when the consequences of deciding that areas of a site are hazardous—when in reality they are not—are relatively minor in early phases of site assessment). In this case, the site manager may make a decision during this stage of

the investigation by using a moderate amount of data, analyzed using a field screening analytical method, and only using a limited number of confirmatory analyses.

Conversely, if the consequences of decision errors are severe, the site manager will want to develop a data collection design that exercises more control over sampling design and measurement error. For example, during the cleanup attainment evaluation phase, deciding that a site is not hazardous when it truly is may have serious consequences because the site may pose a risk to human health and to the environment. Therefore, the decision made during this phase of the assessment process may need to be supported by a large amount of data, and analyzed using very precise and accurate analytical methods.

A site manager should balance the consequences of a decision error against the cost of limiting the possibility of this error. It may be necessary to iterate between Step 6 and Step 7 several times before this balance between limits on decision errors and costs of data collection design is achieved. This is not an easy part of the DQO Process. The balancing of the risk of incorrect decision with potential consequences should be fully explored by the planning team. Resorting to arbitrary values such as "false rejection = 0.05, false acceptance = 0.20" is not recommended; the circumstances of the investigation may allow for a less stringent choice, or possibly a more stringent requirement. In the early stages of DQO development, it is recommended that a very stringent choice be made and the consequences of that choice be investigated by using the DEFT software (EPA, 1994c).

6.2 ACTIVITIES

The following subsections describe the process of establishing decision performance criteria. The combined information from these activities is graphically displayed as Decision Performance Goal Diagrams (DPGDs) in Figures 5 and 6 or charted in decision error limits tables in Tables 7 and 8. Both of these methods illustrate the site manager's tolerable risk of decision errors.

How to Read a Decision Performance Goal Diagram: Figures 5 and 6 show in graphical form some key outputs of Step 6 of the DQO Process. The full meaning and interpretation of a Decision Performance Goal Diagram (DPGD) should be clear after reading the rest of section 6.2. As the explanation progresses, it may be helpful to keep in mind that the DPGD represents a set of "what if?" conditions in the following sense. A decision maker asks, "*what if* the *true* concentration of contaminants was *this* high and how strong is my aversion to having the data mislead me into taking the wrong action?" The true concentration is represented on the horizontal axis. The decision maker's aversion to taking a wrong action is expressed as tolerable probabilities of committing a decision error, which are indicated along the vertical axis. The action level defines the true concentration above which some action should be taken (such as further investigation or remediation).

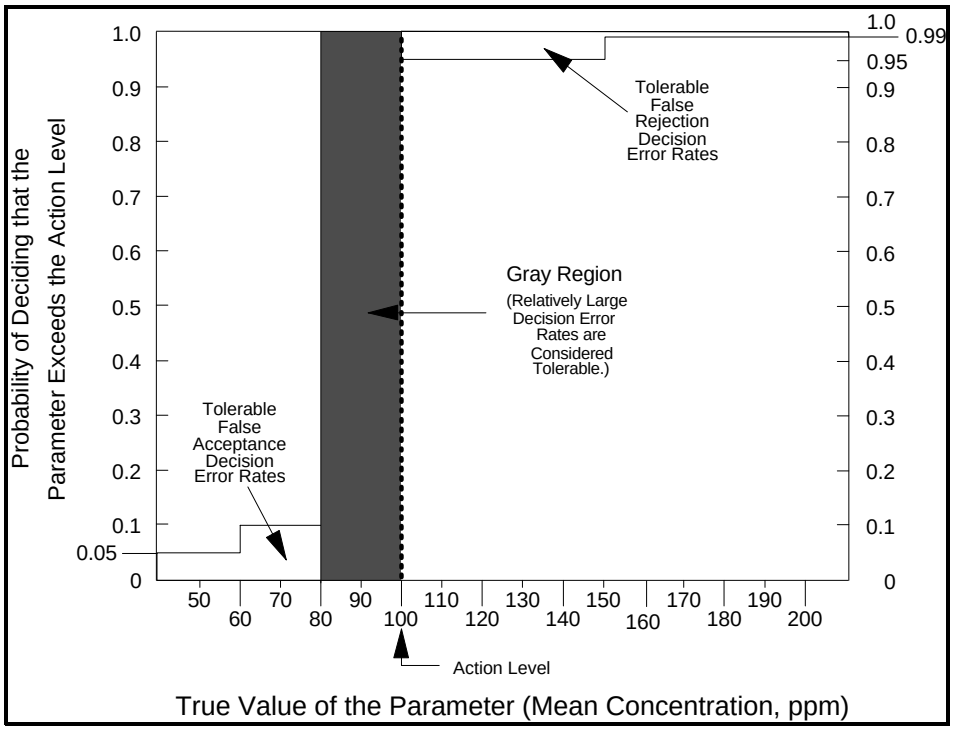


Figure 5. An Example of a Decision Performance Goal Diagram – Baseline Condition: Parameter <u>Exceeds</u> Action Level

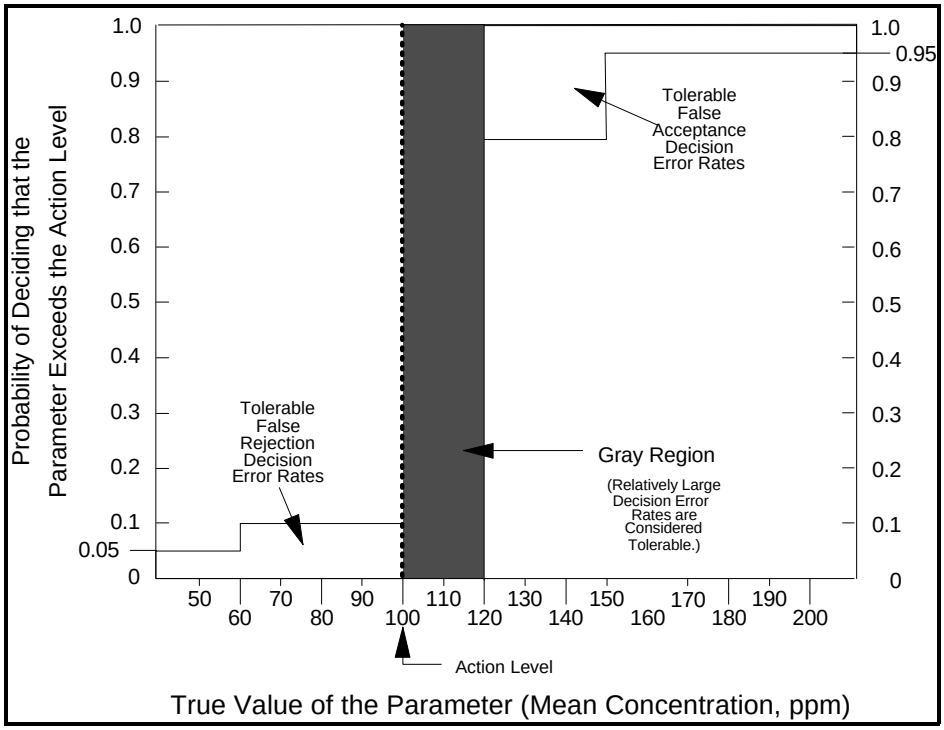


Figure 6. An Example of a Decision Performance Goal Diagram – Baseline Condition Parameter <u>Is Less Than</u> Action Level

Table 7. Decision Error Limits Table Corresponding to Figure 5

True Concentration	Correct Decision	Tolerable Probability of Making an Incorrect Decision
0 to 60 ppm	does not exceed action level	5%
60 to 80 ppm	does not exceed action level	10%
80 to 100 ppm	does not exceed action level	gray region—no probability specified
100 to 150 ppm	exceeds action level	10%
greater than 150 ppm	exceeds action level	1%

Table 8. Decision Error Limits Table Corresponding to Figure 6

True Concentration	Correct Decision	Tolerable Probability of Making an Incorrect Decision
0 to 60 ppm	does not exceed action level	5%
60 to 100 ppm	does not exceed action level	10%
100 to 120 ppm	exceeds action level	gray region—no probability specified
120 to 150 ppm	exceeds action level	20%
greater than 150 ppm	exceeds action level	5%

6.2.1 Determine the Possible Range of the Parameter of Interest

The planning team should establish the possible range of the parameter of interest by estimating its upper and lower bounds based on currently available information and professional judgment. This helps focus the process of defining probability limits on decision errors on only

the relevant values of the parameter. The team may use historical data, including analytical data (if they are available) as a starting point for defining the possible range of the parameter of interest. The team also should ensure that the range is sufficiently wide to account for uncertainties or gaps in the information used to set the range. For example, if the parameter of interest is a mean, the range may be defined using the lowest and highest concentrations at which the contaminant is thought to exist at the site.

Example: The range of the population mean shown in Figures 5 and 6 is between 0 and 210 ppm. Note that for purposes of interpreting a DPGD, the concentration values on the horizontal axis represent true values of the parameter of interest, not the estimated value from the data.

6.2.2 Define Both Types of Decision Errors, Identify Their Potential Consequences and Select the Baseline Condition

The planning team should designate the areas above and below the action level as the range where the two types of decision errors may occur. Next, the team should define the baseline condition (null hypothesis) based on the relative consequences of the decision errors. This activity has four steps:

(1) **Define both types of decision errors and establish the "true state of nature" for each decision error.** The team should state both decision errors in terms of the parameter of interest, the action level, and the alternative actions. An example of a decision error is to "decide that the true mean concentration of site-related contaminants exceeds the action level and remediation is necessary when in fact the mean concentration of site-related contaminants does not exceed the action level and remediation is not necessary."

The "true state of nature" is the actual condition of the parameter in the media but unknown to the decision maker. Each decision error consists of two parts: the true state of nature and the conclusion the decision maker reaches. For example, the true mean concentration of site-related contaminants does not exceed the action level (the "true state of nature"); however, the site manager has determined from the data that the mean concentration of site-related contaminants exceeds the action level (the conclusion reached by the decision maker).

(2) **Specify and evaluate the potential consequences of each decision error.** The team should consider the consequences of making each decision error. For example, potential consequences of incorrectly deciding that the parameter is below the action level (when in fact it is above the action level) include potential threats to human health and the environment. Conversely, potential consequences of incorrectly deciding that the value of the parameter of interest is above the

action level (when in fact it does not exceed the action level) include spending unnecessary resources to study further and/or possibly to remediate an uncontaminated site.

The team should evaluate the potential consequences of decision errors at several points within the false rejection and false acceptance ranges. For example, the consequences of a decision error when the true parameter value is only just 10% above the action level may be minimal because it may cause only a moderate increase in the risk to human health. Conversely, the consequences of a decision error when the true parameter is an order of magnitude above the action level may be severe because it could significantly increase the risk to human health and threaten the local ecosystem.

(3) **Establish which decision error has more severe consequences near the action level.** The site manager should use the evaluation of the potential consequences of the decision errors to establish which decision error has the more severe consequences near the action level. For example, the site manager would judge the threat of health effects from a site contaminated with acutely hazardous waste against spending unnecessary resources to remediate a clean site.

(4) **Define the baseline condition (null hypothesis) and the alternative condition (alternative hypothesis), and assign the terms "false rejection" and "false acceptance" to the appropriate decision error.** The baseline assumption is the one that will be kept until overwhelming evidence (in the form of data to be collected at the site) is presented to make the site manager reject the baseline assumption in favor of the alternative. One rationale is to set the baseline condition equal to the true state of nature that exists when the more severe error occurs, therefore guarding against the occurrence of this error because the baseline assumption will only be abandoned with reluctance (i.e., weight of the data indicating it should be wrong). A false rejection decision error corresponds to the more severe decision error, and a false acceptance decision error corresponds to the less severe decision error. Note that under some RCRA regulations the choice of baseline and alternative have already been set. For example, in a delisting petition, the baseline condition is that the waste is hazardous, and the alternative is that it is not hazardous. This means that the petitioner has to present overwhelming evidence (data) to show that the baseline is incorrect.

Example. The action level has been set at 100 ppm in Figures 5 and 6. (Note that the action level is represented by a vertical dashed line at 100 ppm.) Figure 5 shows the case where a site manager considers the more severe decision error to occur above the action level. Figure 6 shows the reverse, the case where the site manager considers the more severe decision error to occur below the action level. As the hypothesis test for the second case is the reverse of the first case, the false rejection and false acceptance errors are reversed (on opposite sides of the action level). For illustrative purposes, this chapter focuses on the first case, shown in Figure 5.

6.2.3 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 site manager 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 a site manager 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 width of the gray region may be wide during early phases of the site assessment process, where further evaluation of the site can identify if the parameter of interest is slightly less than the action level. Similarly, during a cleanup attainment evaluation phase, the width of the gray region may also be wide as use of a wide gray region will usually yield conclusive evidence of a successful remediation. However, if the site manager believes that the cleanup process has only remediated to the extent that the parameter is close to the action level, a narrow gray region will be necessary to detect successful remediation. In general, the narrower the gray region, the greater the number of samples needed to meet the criteria.

Example. Consider the DPGD Shown as Figure 5. Notice that the site manager has located the action level at 100 ppm and edge of the gray area at 80 ppm. This implies that when the sample data mean is less than 80 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.

6.2.4 Assign Probability Values to Points Above and Below the Action Level that Reflect the Tolerable Probability for the Occurrence of Decision Errors

A decision error limit is the probability that a decision error may occur for a specific value of the parameter of interest. This probability is an expression of the decision maker's tolerance for uncertainty but does not imply that a decision error will occur. Instead it is only a measure of the risk a decision-maker is willing to assume.

At a minimum, the site manager should specify a false rejection decision error limit at the action level and a false acceptance decision error limit at the other end of the gray region based on the consequences of the respective errors. Severe consequences (such as extreme risks to human health) should have stringent limits (small probabilities), whereas moderate consequences may have less stringent limits. In general, the tolerable limits for making a decision error should decrease as the consequences of a decision error become more severe farther away from the action level.

The most stringent limits on decision errors that are typically encountered for environmental data are 0.01 (1%) for both the false rejection and false acceptance decision errors. This guidance recommends using 0.01 as the starting point for setting decision error rates.[2] If the consequences of a decision error are not severe enough to warrant this stringent decision error limit, this value may be relaxed (a larger probability may be selected). However, if this limit is relaxed from a value of 0.01 for either the decision error rate at the action level or the other bound of the gray region, the planning team should document the rationale for relaxing the decision error rate. This rationale may include regulatory guidelines; potential impacts on cost, human health, and ecological conditions; and sociopolitical consequences.

6.3 OUTPUTS

The outputs from this step are the site manager's tolerable decision limits based on a consideration of the consequences of making an incorrect decision. These limits on decision errors can be expressed in a decision error limits table (Tables 7 and 8) or in a DPGD as illustrated in Figures 5 and 6.

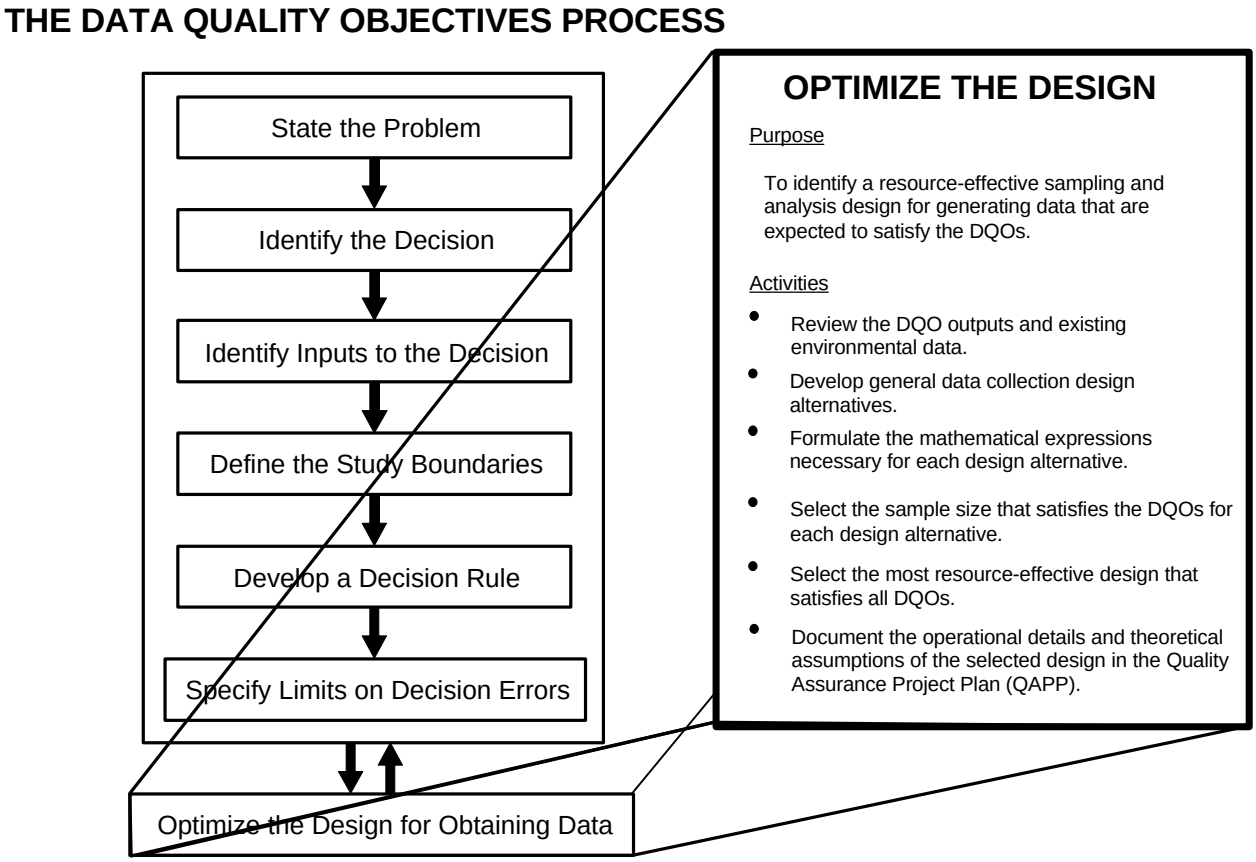
---

[6]The value of 0.01 should not be considered a prescriptive value for setting decision error rates, nor should it be considered as EPA policy to encourage the use of any particular decision error rate.

Example. It is often useful to summarize the decision error limits in either a table or a graph. Figure 5 and Table 7 show that from the action level to a true value of 135 ppm for the parameter of interest, the site manager will tolerate a 10 percent chance of deciding that the true value is below the action level, based on field investigation data. If the true value is greater than 135 ppm, the site manager will tolerate only a 1 percent chance of deciding the true value is really below the action level. Below the action level, from 60 to 80 ppm the site manager will tolerate deciding the true value is above the action level 10 percent of the time, and between 40 and 60 ppm the site manager will allow a false acceptance decision error rate of 5 percent. These probabilities represent the risk to the site manager of making an incorrect decision.

CHAPTER 7

STEP 7: OPTIMIZE THE DESIGN FOR OBTAINING DATA



7.1 BACKGROUND

The purpose of this step is to identify a resource-effective field investigation sampling design that generates data that are expected to satisfy the site manager's decision performance criteria, as specified in the preceding steps of the DQO Process. To develop the optimal design for this study, it may be necessary to work through this step more than once after revisiting previous steps of the DQO Process. The output of this step is the sampling design that will guide development of QA project documentation, such as the field sampling and analysis plan and the QA Project Plan required for EPA investigations.

This step provides a general description of the activities necessary to generate and select data collection designs that satisfy decision performance criteria defined in Step 6, "Specify Tolerable Limits on Decision Errors." In addition, it contains information about how DQO outputs from the previous six steps of the DQO Process are used in developing a statistical

design. However, this document does not give detailed guidance on the mathematical procedures involved in developing a statistical data collection design. Investigators may refer to Cochran (1977) or Thompson (1992) for theoretical discussions, Chapters 2 through 10 of Gilbert (1987), and *Methods for Evaluation of the Attainment of Cleanup Standards: Volume 1* (EPA, 1989a) for more information. It should be stressed that if critical design assumptions are seriously violated, the data may become unusable for the specified purpose.

For most field investigations, a probabilistic sampling approach will be necessary to have a scientific basis for extrapolating results from a set of samples to the entire site or large areas of the site. All probability sampling designs have an element of randomization, which allows probability statements to be made about the quality of estimates derived from the data. Therefore, probability samples are useful for testing hypotheses about whether a site is contaminated, what the level of contamination is, and what other problems common to hazardous waste sites have occurred. By combining an effective probabilistic data collection design with a statistical hypothesis test, the decision maker will be able to optimize resources such as funding, personnel, and time while still meeting DQOs. There are many different probability sampling designs, each with advantages and disadvantages. A few of the most basic sampling designs are described in Table 9; other probability designs, such as rank set sampling and search sampling, are beyond the scope of this guidance.

A nonprobabilistic sampling (judgmental sampling) design is developed when the site manager (or technical expert) selects the specific sampling locations based on the investigator's experience and expert knowledge of the site. Typically, this is useful to confirm the existence of contamination at specific locations, based on visual or historical information. Judgmental samples can be used subjectively to provide information about specific areas of the site, often useful during the preliminary assessment and site investigation stages—provided there is substantial information on the contamination sources and history.

However, when nonprobabilistic sampling approaches are used, quantitative statements about data quality are limited only to the measurement error component of total study error and the results cannot be extrapolated to the entire site unless the data are being used to support explicit (usually deterministic) scientific models, such as ground water contaminant fate and transport.

If a judgmental data collection design is chosen, it is important to implement and document the applicable activities of this DQO step. This approach will help the planning team document the reasons for selecting a nonprobabilistic sampling scheme, the reasons for selecting specific sampling locations, and the expected performance of the data collection design with respect to qualitative DQOs only. If the site manager wishes to draw conclusions about areas of the site beyond the exact locations where samples were taken or if statistically defensible conclusions are desired, then a probabilistic approach should be used.

Table 9. Probabilistic Sampling Designs

Simple Random Sampling —The basic probability sample is the simple random sample (SRS). With SRS, every possible sampling point has an equal probability of being selected, and each sample point is selected independently from all other sample points. *Pros*: SRS is appropriate when little information is available for a site, and the population does not contain any trends. *Cons*: If some information is available, SRS may not be the most cost-effective (efficient) sampling design.
Systematic Sampling — Systematic sampling achieves a more uniform spread of sampling points than SRS by selecting sample locations using a spatial grid, such as a square, rectangle, or triangle, in two or three dimensions. Pros: Sampling locations are located at equally spaced points so they may be easier to locate in the field than simple random samples. Cons: A systematic sample should not be used if the contamination exhibits any cyclical patterns.
Stratification — Stratified random sampling is used to improve the precision of a sampling design. For stratified sampling, the study area is split into two or more nonoverlapping strata (subareas) where physical samples within a stratum are more similar to each other than to samples from other strata. Sampling depth, concentration level, previous cleanup attempts, and confounding contaminants can be used as the basis for creating strata. Stratification is an accepted way to incorporate prior knowledge and professional judgment into a probabilistic sampling design. Once the strata have been defined, each stratum is then sampled separately using one of the simple methods (e.g., SRS). *Pros*: A stratified sample can be more cost-effective and can be used to ensure that important areas of the site are represented in the sample. In addition, parameter estimates can be developed for each stratum. *Cons*: Analysis of the data is more complicated than for other sampling designs.
Composite Sampling — Composite sampling is used to estimate the population mean when chemical analysis costs are high compared to sampling costs or if the between-sample variability is much larger than analytical variability. Composite sampling involves physically mixing two or more samples before analysis. This method should be used in conjunction with a sample design in order to determine sample locations (e.g., SRS with compositing). *Pros*: Composite sampling can be a cost-effective way to select a large number of sampling units and provide better coverage of the site without analyzing each unit. It also is useful if the samples will be used as a screening device. *Cons*: Composite sampling should not be used when information about extreme values or variability is required or when samples are changed by the mixing process (e.g., volatile chemicals).

7.2 ACTIVITIES

7.2.1 Review the DQO Outputs and Existing Environmental Data

The outputs from the previous steps of the DQO Process provide a succinct collection of information that is used to develop the data collection design in the following ways:

- the inputs, boundaries, and decision rule are used in determining the type, location, and timing of samples; and

- the limits on decision errors provide crucial information for selecting the number of samples to be collected and the number of analyses per sample.

Information regarding the expected variability of contaminants is necessary for most probabilistic data collection designs, and any existing environmental data from the site (or from similar sites) should be reviewed for potential use in statistical analysis or in defining the boundaries of the study. Information about existing environmental data may have been identified during Step 1, "State the Problem," and Step 3, "Identify the Inputs to the Decision." If no existing data are available, it may be necessary to conduct a limited field investigation to acquire an preliminary estimate of variability.

7.2.2 Develop General Data Collection Design Alternatives

The planning team should develop alternative data collection designs that could generate data needed to test the hypothesis. These alternatives should, at a minimum, include the sample selection technique, the sample type, the sample size, and the number of analyses per sample. To generate alternative designs, the planning team may vary the sampling design, the type of samples collected, the field sampling or analytical methods used, or the number of replicate analyses performed on samples.

It is important not to rule out any alternative field sampling or analytical methods due to preconceptions about whether or not the method is sufficient. It should be remembered that the objective of the statistical design is to limit the total study error, which is a combination of sampling design and measurement error, to tolerable levels so that the site manager's decision performance criteria are satisfied. Designs that balance the number of field samples with the number of laboratory analyses should be considered.

7.2.3 Formulate the Mathematical Expressions Necessary for Each Design Alternative

Two mathematical expressions are necessary for optimizing each data collection design alternative in relation to the decision performance criteria. First, a tentative method for analyzing the resulting data (e.g., a student's *t*-test or a tolerance interval) should be specified, along with any available sample size formulas corresponding to the proposed method. This information will

be used to solve for the minimum sample size that satisfies the decision maker's limits on decision errors. Second, a cost function that relates the total number of samples to the costs of sampling and analysis should be developed. This information will be used to compare the cost-effectiveness of different sampling designs.

Some common data analysis methods and sample size formulas are contained in Table 10 and described in-depth in *Guidance for Data Quality Assessment (QA/G-9)* (EPA, 1996b). The types of tests applied at hazardous waste sites can be broadly classified as one-sample (single-site) tests or two-sample (double-site) tests. In one-sample cases, data from a site are compared with an absolute criterion such as a regulatory threshold or an ARAR. In this case, the parameter of interest is usually a mean, median, percentile, or proportion of contamination levels within each scale of decision making, such as an EU. In the two-sample cases, data from a site are compared with data from another site or background area. In this case, the parameter of interest is usually the difference between the two means, two medians, two proportions, or two percentiles, and the action level is often zero. If two independent random samples are taken at the same site at two different times, such as before and after some remediation activity, then the first set of measurements can be interpreted as if for site A, and the second set of measurements can be interpreted as if for site B.

7.2.4 Select the Sample Size That Satisfies the DQOs for Each Design Alternative

The planning team should calculate the sample size for each data collection design alternative. If none of the data collection designs satisfies all of the decision performance criteria (including cost), the planning team may need to:

- increase the tolerable limits on decision errors;
- increase the width of the gray region;
- increase funding for sampling and analysis;
- change the boundaries (it may be possible to reduce sampling and analysis costs by changing or eliminating subgroups that will require separate decisions); or
- relax other project constraints.

To assist the team in generating their development of alternative designs, EPA has developed the software, *Data Quality Objectives Decision Error Feasibility Trials (QA/G-4D) (DEFT)* (EPA, 1994c). DEFT is a personal computer software package developed to assist the site manager and planning team in evaluating whether the DQOs are feasible before the development of the final data collection design is started. To do this, DEFT software uses the DQO outputs generated in Steps 1 through 6 of the DQO Process to evaluate several basic data collection designs, including simple random sampling, simple random sampling with composite

Table 10. Common Sample Size Formulas

Statistical Test	Parameter of Interest and Baseline Conditions	Sample Size Formula
One-Sample *t*-test	Parameter of Interest: Mean Baseline Conditions: Mean ≤ AL Mean ≥ AL	$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2 s^2}{\Delta^2} + \frac{z_{1-\alpha}^2}{2}$
One-Sample Test for a Proportion	Parameter of Interest: Proportion Percentile Baseline Conditions: Proportion (Percentile) ≤ AL Proportion (Percentile) ≥ AL	$n = \left[\frac{z_{1-\alpha}\sqrt{AL(1-AL)} + z_{1-\beta}\sqrt{GR(1-GR)}}{\Delta}\right]^2$
Two-sample *t*-test[1]	Parameter of Interest: Difference between two means Baseline Conditions mean(site1) - mean(site2) ≤ 0 mean(site1) - mean(site2) ≥ 0	$n = 2\frac{(z_{1-\alpha}+z_{1-\beta})^2 s^2}{\Delta^2} + \frac{z_{1-\alpha}^2}{4}$
Two-Sample Test for Proportions[1]	Parameter of Interest: Difference between two percentiles Baseline Conditions: Site 1 percentile is less than or equal to Site 2 percentile	$n = \frac{2(z_{1-\alpha}+z_{1-\beta})^2 \bar{P}(1-\bar{P})}{(P_2-P_1)^2}$ where $\bar{P} = \frac{(P_1+P_2)}{2}$

Notation: AL = Action Level
GR = other bound of the gray region
α = the false rejection error rate at the action level
β = the false acceptance error rate at the other bound of the gray region
P = proportion
s = estimate of the standard deviation
Δ = width of gray region

samples, and stratified random sampling. The software then estimates the number of samples and the associated cost required to meet DQOs of each data collection design under consideration.

If the DQOs are not feasible, DEFT software allows the site manager to relax some of the DQOs until a feasible alternative is achieved. The software allows the user to change the action level, the false rejection error rates, the false acceptance error rates, the gray region, the estimate of the standard deviation, and the sample collection and analysis costs. For each change, the software computes a new sample size and total cost, which the site manager can evaluate. If the DQOs are feasible but do not take full advantage of the sampling and analysis budget, the site manager can use DEFT software to specify more stringent DQOs.

7.2.5 Select the Most Resource-Effective Design that Satisfies all DQOs

The planning team should perform a sensitivity analysis on the alternative designs to see how each design performs when the assumptions are changed and to view the impact on costs and resources. Typically, this analysis involves changing certain design parameters within some reasonable range, and seeing how each of these changes influences the ability of the design to achieve expected decision error limits. For example, if the contaminant variability is higher or lower than assumed for the design, what happens to the design performance? Or, if the final remediation level is more or less stringent than the assumed action level, what happens to the decision performance goals?

A performance curve is extremely useful in investigating the expected performance of alternative designs to determine if they are likely to satisfy the DQOs established and to compare several different alternative designs. A performance curve, which is similar in concept to a statistical power curve, represents the probability of deciding that the parameter of interest is greater than the action level over the range of possible population parameters. When no error is associated with a decision, there are two possibilities: the ideal performance curve is equal to zero if the parameter is less than the action level, and equal to one if the parameter is above the action level. In other words, in an ideal world, the risk of making any decision errors would be zero and the gray region would simply be the action level. However, because decisions are based on imperfect data, it is impossible to achieve this ideal power function. Instead, the performance curve will most likely yield values that are small below the action level and large above the action level. Figure 7 shows the difference between the graphs of an ideal performance curve and a realistic performance curve function. A design that produces a very steep performance curve (i.e., closer to the ideal) is preferred over one that is relatively flat, all other things (such as cost) being equal. Figure 8 shows a performance curve overlaid on a Decision Performance Goal Diagram.

7.2.6 Document the Operational Details and Theoretical Assumptions of the Selected Design in the Quality Assurance Project Plan

Once the final data collection design has been selected, it is important to ensure the design is properly documented. This improves efficiency and effectiveness of later stages of the data

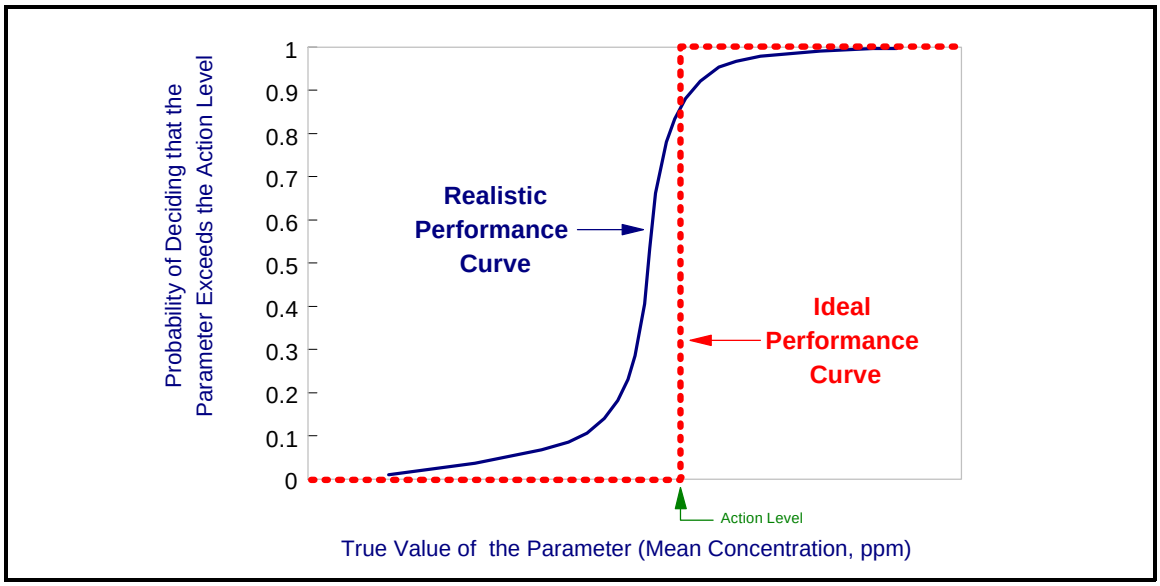


Figure 7. Ideal Versus Realistic Performance Curve

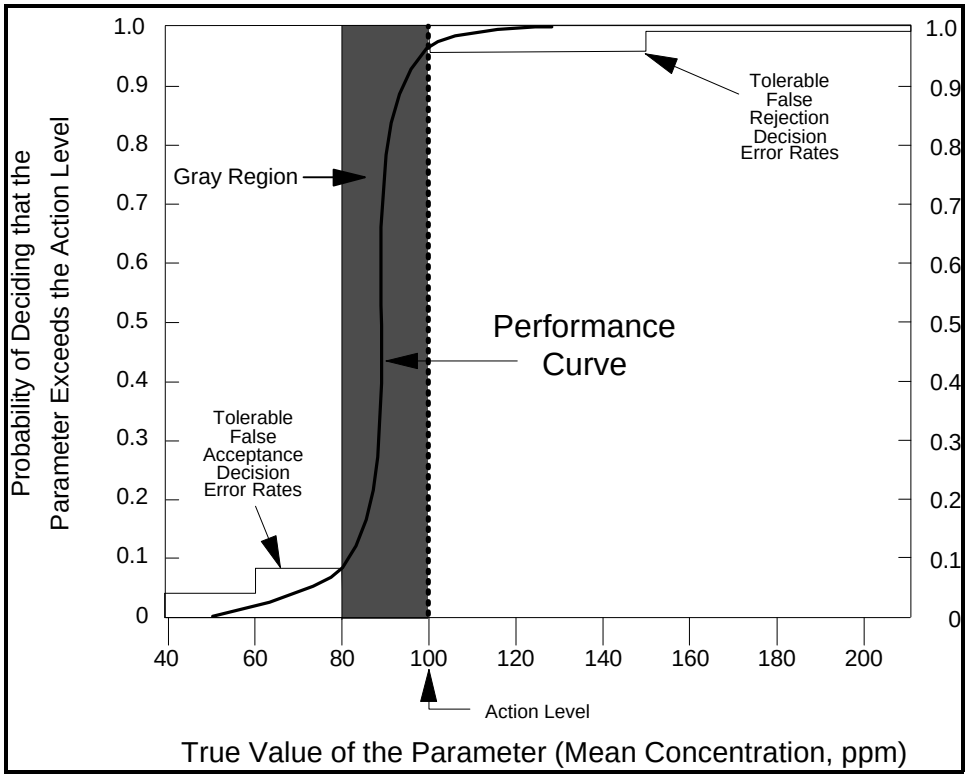


Figure 8. An Example of a Performance Curve Overlaid on a Decision Performance Goal Diagram (Baseline Condition: Parameter Exceeds Action Level)

collection and analysis process, such as the development of field sampling procedures, QC procedures, and statistical procedures for data analysis. The key to successful design documentation is in drawing the link between the statistical assumptions on which the design is based and the practical activities that ensure these assumptions generally hold true.

For EPA programs, the operational requirements for implementing the data collection design are documented in the Field Sampling Plan and the QA Project Plan. Design elements that should be documented include:

- sample size;
- sample type (e.g., composite vs. grab samples);
- general collection techniques (e.g., split spoon vs. core drill, or activated charcoal media vs. evacuated canister);
- sample support (i.e., the amount of material to be collected for each sample);
- sample locations (surface coordinates and depth) and how locations were selected;
- timing issues for sample collection, handling, and analysis;
- analytical methods (or performance standards); and
- QA and QC protocols.

Note that proper documentation of the model and assumptions used for collecting data is essential to maintain the overall validity of the study in the face of unavoidable deviations from the original design. In some cases, the QA Project Plan can be used instead of a Field Sampling Plan but this will depend on the decision of the site manager.

7.3 OUTPUTS

The outputs for this step include the optimal (most resource-effective) data collection design for the field investigation, along with documentation of the key assumptions underlying the design.

CHAPTER 8

BEYOND THE DQO PROCESS: QUALITY ASSURANCE PROJECT PLANS AND DATA QUALITY ASSESSMENT

8.1 OVERVIEW

This chapter outlines some important quality management steps and actions that occur after the DQO Process has been completed.

8.2 THE PROJECT LIFE CYCLE

A project's life cycle consists of three principal phases: planning, implementation, and assessment . Each of these three phases demand attention to quality assurance issues and these issues are illustrated in Figure 9. This document focuses on just the planning phase.

8.2.1 Planning

During the planning stage, investigators specify the intended use of the data to be collected and plan the management and technical activities (such as sampling) that are needed to generate the data. The DQO Process is the foundation for the planning stage and is supported by

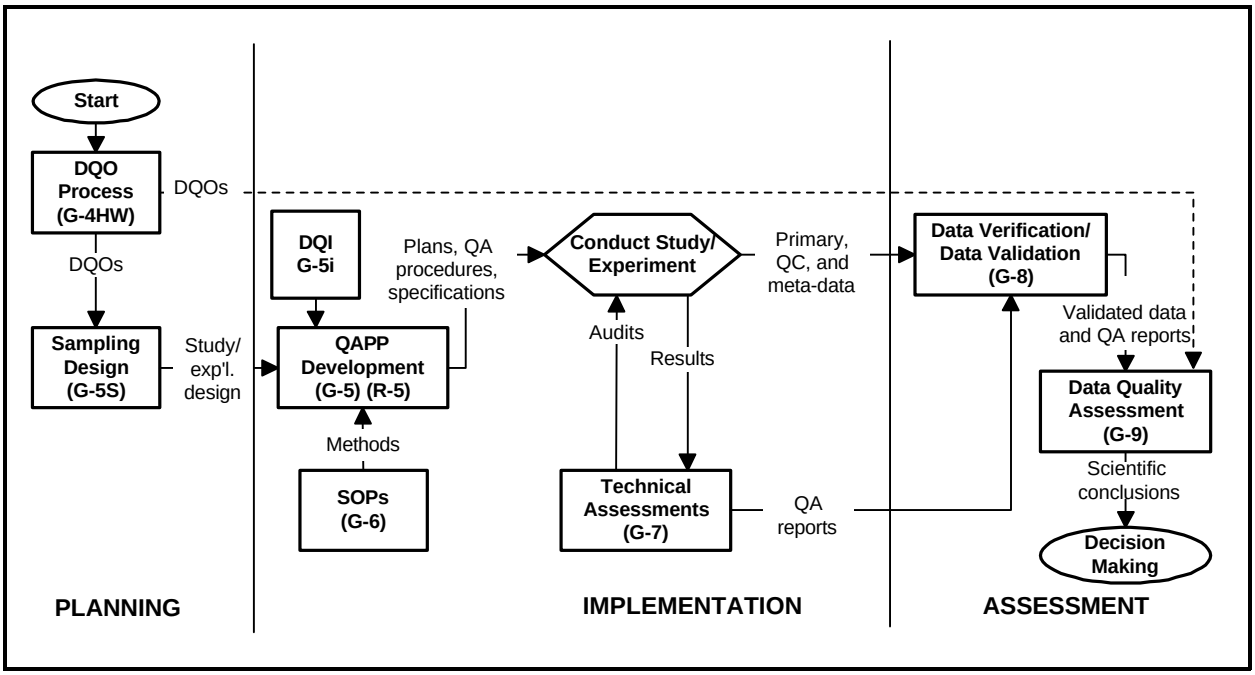


Figure 9. The DQO Process is the Initial Component of the Project Level of EPA's Quality System *(For each component, the corresponding Quality Series document is denoted.)*

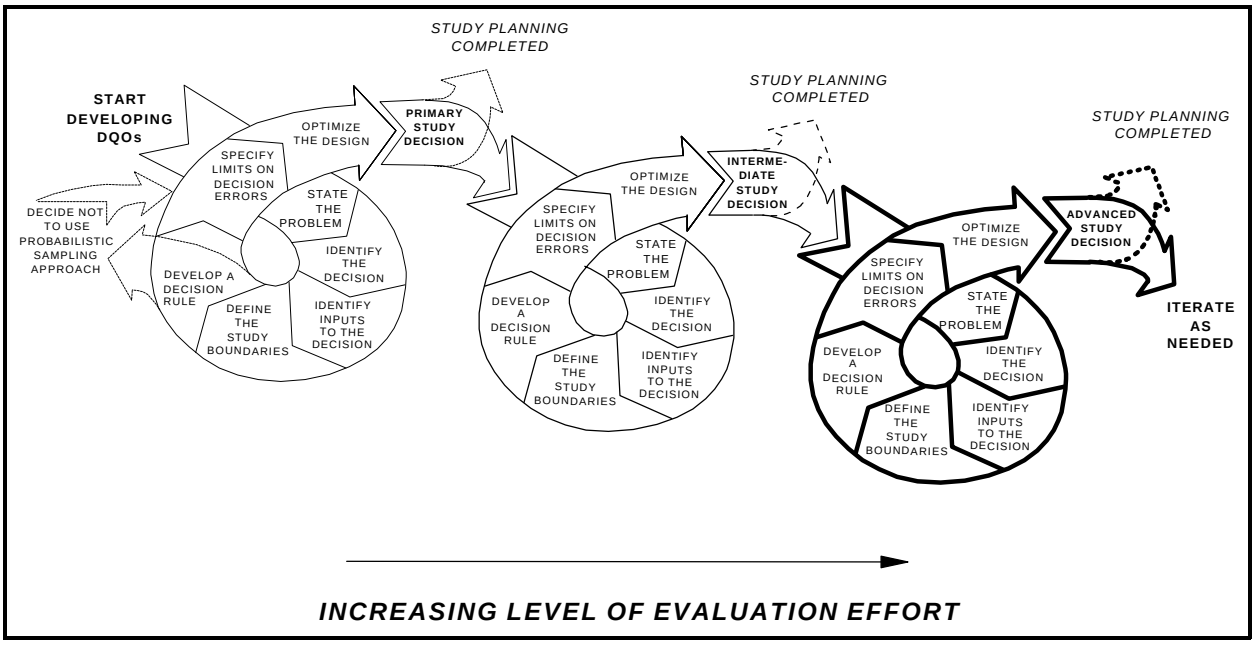


Figure 10. The Iterative Nature of the DQO Process

a sampling design, the generation of appropriate data quality indicators, standard operating procedures, and finally the mandatory QA Project Plan. The DQO Process is iterative (Figure 10) and is allowed to terminate when the DQO outputs are acceptable to the decision maker with respect to potential decision error rates and expenditure of resources.

8.2.2 Implementation

During the implementation phase of the Data Life Cycle, investigators collect and analyze samples according to the specifications of the QA Project Plan and the field sampling and analysis plan. QA and QC protocols such as technical systems audits and performance evaluations are conducted to ensure that data collection activities are conducted correctly and in accordance with the QA Project Plan. In Superfund Remedial Investigations (RIs), the sampling design and the other DQO outputs are used to develop the QA Project Plan and the FSP, which in turn are combined to create the SAP. The SAP provides detailed site-specific objectives, QA and QC specifications, and procedures for conducting a successful field investigation that are intended to produce data of the quality needed to satisfy the site manager's decision performance criteria. In the RCRA Corrective Action Program, the DQO Process can be used to prepare for RFIs. Both the QA Project Plan and the sampling design are then combined to create the RFI Workplan.

A QA Project Plan is composed of up to 24 elements grouped into four classes-project management, measurement/data acquisition, assessment/oversight, and data validation and usability (Table 11). Not all elements need to be addressed for every project. However, other projects may require additional information that is not contained in the 24 elements. The final decision on what elements need to be addressed is made by the overseeing or sponsoring EPA

Table 11. QA Project Plan Elements

A. Project Management

A1 Title and Approval Sheet
A2 Table of Contents
A3 Distribution List
A4 Project/Task Organization
A5 Problem Definition/Background
A6 Project/Task Description
A7 Quality Objectives and Criteria for Measurement Data
A8 Special Training Certification
A9 Documents and Records

B. Measurement/Data Acquisition

B1 Sampling Process Design
B2 Sampling Methods (Experimental Design)
B3 Sample Handling and Custody
B4 Analytical Methods
B5 Quality Control
B6 Instrument/Equipment Testing, Inspection, and Maintenance
B7 Instrument/Equipment Calibration and Frequency
B8 Inspection/Acceptance Requirements for Supplies and Consumables
B9 Non-Direct Measurements
B10 Data Management

C. Assessment/Oversight

C1 Assessments and Response Actions
C2 Reports to Management

D. Data Validation and Usability

D1 Data Review, Verification, and Validation
D2 Verification and Validation Methods
D3 Reconciliation with User Requirements

organization. No environmental data collection or use may occur without an EPA-approved QA Project Plan in place except under special conditions.

Class A: Project Management. This class of QA Project Plan elements addresses project management, project history and objectives, and roles and responsibilities of the participants. Class A elements help ensure that project goals are clearly stated, that participants understand the project goals and approach, and that the planning process in documented.

Class B: Measurement/Data Acquisition. Class B elements cover all aspects of the measurement system design and implementation as well as ensure that appropriate methods for sampling, analysis, data handling, and QC are employed and documented. Goals for data quality are specified in this class.

Class C: Assessment/Oversight. This class will help to ensure that the QA Project Plan is implemented as prescribed. Class C elements address activities for assessing the effectiveness of project implementation and associated QA and QC.

Class D: Data Validation and Usability. This class of elements helps to ensure that data meet the specified criteria. Class D elements address QA activities that occur after data collection is complete.

Guidance documents useful to ensure the successful implementation of the project are:

- *Guidance for QA Project Plans (QA/G-5)* (EPA, 1998c). Guidance on the construction of the mandatory plan for data collection that describes the necessary QA and QC activities that should be implemented in order to ensure the data will be sufficient to meet the intended DQOs.

- *Guidance for the Preparation of SOPs for Quality-Related Documents (QA/G-6)* (EPA, 1995b). A general description of the format for SOP documents.

- *Guidance on Technical Assessments for Environmental Data Operations (QA/G-7)* (EPA, 2000). This document describes various kinds of assessments such as technical systems audits that are important to ensure data and information are being produced according to the QA Project Plan.

8.2.3 Assessment

During the assessment phase, data are verified and validated in accordance with the QA Project Plan, and a DQA is performed to determine if the DQOs have been satisfied. DQA is a scientific and statistical evaluation to determine whether environmental data are of the right type, quality, and quantity to support Agency decisions. DQA consists of five steps that parallel the activities of a statistician analyzing a data set for the first time. However, it makes use of statistical and graphical tools that even nonstatisticians can apply to data sets.

DQA is built on a fundamental premise: data quality, as a concept, is meaningful only when it relates to the intended use of the data. Data quality does not exist without some frame of reference; one must know the context in which the data will be used in order to establish a yardstick for judging whether or not the data set is adequate.

By performing DQA, environmental scientists and managers can answer two fundamental questions: (1) Can the decision (or estimate) be made with the desired confidence, given the quality of the data set? and (2) How well can the sampling design used to collect the data set be expected to perform in other data collection events under difference conditions? The first question addresses the data user's immediate needs. For example, if the data provide evidence strongly in favor of one course of action over another, then the decision maker can proceed

knowing that the decision will be supported by unambiguous data. If the data do not show sufficient evidence to favor one alternative, then the data analysis alerts the decision maker to this uncertainty. The second question addresses the data user's future needs. Often, investigators decide to use a certain sampling design at a location different from that for which it was first designed. In these cases, they should determine how well the design is expected to perform given that the outcomes and environmental conditions will differ from those of the original event. By estimating the outcomes before the sampling design is implemented, investigators can make any necessary modifications and thus prevent costly additional follow-up rounds of sampling to supplement inadequate data. DQA (see Figure 11) involves the application of statistical tools to determine whether the variability and bias in the data are small enough to allow the site manager to use the data to support the decision with acceptable confidence.

Guidance for the assessment phase includes:

- *Guidance for Data Quality Assessment (QA/G-9)* (EPA, 1996b). The scientific and statistical process that determines whether the data meet the desired DQO. These practical methods for data analysis are supplemented by software *Data Quality Evaluation Statistical Toolbox (QA/G-9D) (DataQUEST)* (EPA, 1997).

To conclude the assessment phase, it is necessary to document all the relevant information collected over all phases of the project's life cycle. The conclusion from a DQA must be presented in a fashion that facilitates the comprehension of the important points. Care should be taken to explaining statistical nomenclature and avoid use of statistical jargon whenever possible.

8.2.4 Beyond Data Quality Assessment

The ultimate goal of the DQO Process is to collect data of the right type, quality, and quantity to support defensible site decisions; DQA is the final step in ensuring this goal has been reached. One aspect of the entire process that should not be overlooked is the documentation of results obtained during DQA because future studies may have need of important statistical information derived during the investigation of data to confirm their conformance to the planned DQO. The importance of maintaining a unified documentation throughout the entire life cycle of a project cannot be under estimated.

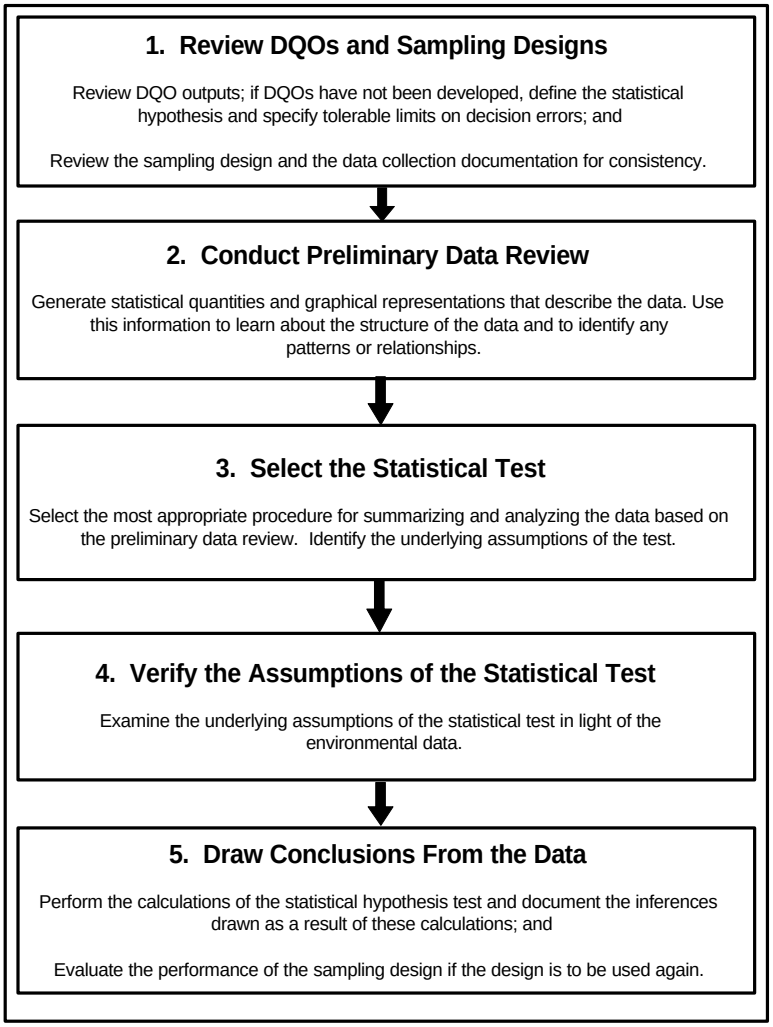


Figure 11. Data Quality Assessment

REFERENCES

American Society for Testing and Materials (ASTM). January 1996. *Standard Practice for Generation of Environmental Data Related to Waste Management Activities: Development of Data Quality Objectives.* D5792-95.

Berthouex, P.M., and L.C. Brown. 1994. *Statistics for Environmental Engineers.* Boca Raton: Lewis.

Cochran, W. 1977. *Sampling Techniques.* New York: John Wiley & Sons.

Department of Energy (DOE). December 1993. Module 7, "Streamlined Approach for Environmental Restoration," in *Remedial Investigation Feasibility Study (RI/FS) Process, Elements and Technical Guidance.* EH 94007658.

Federal Register Vol. 55, No. 145, July 27, 1990

Federal Register Vol. 61, No. 85, May 1, 1996.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring.* New York: Van Nostrand Reinhold (now known as John Wiley & Sons).

Guenther, W.C. 1977. Sample size formulas for normal theory T test. *The American Statistician* **35**, No.4.

Guenther, W.C. 1981. *Sampling Inspection in Statistical Quality Control. Griffin's Statistical Monographs and Courses, No. 37.* London: Charles Griffin.

Lehman, E.L. 1975. *Nonparametrics, Statistical Methods Based on Ranks.* New York: McGraw-Hill.

Myers, J.C. 1997. *Geostatistical Error Management.* New York: Van Nostrand Reinhold (now known as John Wiley & Sons).

Ott, W.R. 1995. *Statistics and Data Analysis.* Boca Raton: Lewis.

Smyth, J, D, and R. D. Quinn (1991). *The Observational Approach in Environmental Restoration.* Presented at the 1991 American Society of Civil Engineers, July 8-10, Reno, NV. PNL-SA-18817. Pacific Northwest National Laboratory, Richland, WA.

Terzaghi, K. (1961). Past and future of applied soil mechanics. *J. Boston Soc. Civ Engr* **68**, 110-139.

Thompson, S.K. 1992. *Sampling.* New York: John Wiley & Sons.

U.S. Environmental Protection Agency (EPA). 1986. *RCRA Facility Assessment (RFA) Guidance.*

U. S. Environmental Protection Agency (EPA). October 1988. *Guidance for Conducting Remedial Investigations and Feasibility Studies Under CERCLA.* Office of Emergency and Remedial Response. EPA/540/G-89/004.

U. S. Environmental Protection Agency (EPA). February 1989a. *Methods for Evaluation of the Attainment of Cleanup Standards: Volume 1—Soils and Solid Media.* Office of Policy, Planning, and Evaluation. EPA 230/R-92-14.

U.S. Environmental Protection Agency (EPA). 1989b. *RCRA Facility Investigation (RFI) Guidance. Volume I - IV.*

U.S. Environmental Protection Agency (EPA). 1989c. *Guidance Document on the Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities.*

U. S. Environmental Protection Agency (EPA). 1989d *Risk Assessment Guidance for Superfund: Volume I—Human Health Evaluation Manual, Part A, Interim Final*. Office of Research and Development. EPA/540/l-89/002.

U.S. Environmental Protection Agency (EPA). 1991a. *Guidance for Performing Preliminary Assessments under CERCLA.*

U.S. Environmental Protection Agency (EPA). 1991b. *Considerations in Ground-Water Remediation at Superfund Sites and RCRA Facilities.*

U. S. Environmental Protection Agency (EPA). 1991c. *Risk Assessment Guidance for Superfund: Volume I—Human Health Evaluation Manual, Part B, Development of Risk-Based Preliminary Remediation Goals*. Office of Research and Development. EPA/540/R-92/003.

U. S. Environmental Protection Agency (EPA). 1991d. *Risk Assessment Guidance for Superfund: Volume II—Environmental Evaluation Manual.* Office of Emergency and Remedial Response. EPA 540/1-89/001.

U.S. Environmental Protection Agency (EPA). 1992a. *Guidance for Performing Site Inspections under CERCLA.*

U.S. Environmental Protection Agency (EPA). 1992b. *Methods for Evaluating the Attainment of Cleanup Standards, Volume 2: Ground Water.* EPA 230-R-92-014.

U. S. Environmental Protection Agency (EPA). 1992c. *Guidance for Performing Site Inspections Under CERCLA.* Office of Emergency and Remedial Response. EPA/540/R-92/021.

U.S. Environmental Protection Agency (EPA). 1992d. *Guidance for Evaluating the Technical Impracticability of Ground Water Restoration.*

U. S. Environmental Protection Agency (EPA). 1993. *Data Quality Objectives Process for Superfund: Interim Final Guidance.* Office of Research and Development. EPA 540-R-93-071.

U.S. Environmental Protection Agency (EPA). 1994a. *RCRA Corrective Action Plan.*

U. S. Environmental Protection Agency (EPA). 1994b. *Guidance for the Data Quality Objectives Process (EPA QA/G-4).* Office of Research and Development. EPA/600/R-96/055.

U. S. Environmental Protection Agency (EPA). 1994c. *Data Quality Objectives Decision Error Feasibility Trials (DEFT) Software (EPA QA/G-4D).* Office of Research and Development. EPA/600/R-96/056.

U.S. Environmental Protection Agency (EPA). 1995a. *RCRA Corrective Action Inspection Guidance Manual.*

U. S. Environmental Protection Agency (EPA). 1995b. *Guidance for the Preparation of Standard Operating Procedures (SOPs) for Quality-Related Documents (EPA QA/G-6).* Office of Research and Development. EPA/600/R-96/027.

U. S. Environmental Protection Agency (EPA). 1996a. *Soil Screening Guidance: User's Guide.* Office of Solid Waste and Emergency Response. EPA/540/R-96/0180.

U. S. Environmental Protection Agency (EPA). 1996b. *Guidance for Data Quality Assessment: Practical Methods for Data Analysis (EPA QA/G-9).* Office of Research and Development. EPA/600/R-96/084.

U. S. Environmental Protection Agency (EPA). 1997. *Data Quality Evaluation Statistical Toolbox (DataQUEST) (EPA QA/G-9D).* Office of Research and Development. EPA/600/R-96/085.

U. S. Environmental Protection Agency (EPA). 1998a. *EPA Quality Manual for Environmental Programs.* EPA 5360. Office of Research and Development.

U. S. Environmental Protection Agency (EPA). 1998b. *EPA Order 5360.1, CHG. 1. Policy and Program Requirements for the Mandatory Agency-wide Quality System.*

U. S. Environmental Protection Agency (EPA). 1998c. *Guidance for Quality Assurance Project Plans (EPA QA/G-5).* EPA/600/R-98/018.

U. S. Environmental Protection Agency (EPA). 2000. *Guidance for Technical Assessments, (EPA QA/G-7).* EPA/600/R-99/080.

APPENDIX A

A COMPARISON OF DQO PROCESS DOCUMENTS

The EPA developed the DQO Process as an approach that allows decision makers to specify measures of the quality of their decisions in order to resolve questions on the type, quality, and quantity of data needed to support these decisions. This process represents an evolution from valid concerns about the *quality of data* to concerns about the *quality of decisions* that will be made from the data. Several federal agencies and industry groups have developed their own guidance on implementing the DQO Process to meet their own needs. Although these guidance documents may appear to be different in some respects, the various approaches generally reflect the specific concerns and priorities of the sponsoring organization rather than fundamental differences in philosophy. These approaches are all based on the need to make decisions under uncertain conditions in environmental protection activities, and all have the DQO Process at their core.

This appendix reviews and compares three guidance documents developed by different organizations (two federal agencies and one national standards organization) that include methods labeled the DQO Process. This review identifies similarities and differences among the documents, discusses how the differences may influence users of the documents, and shows that the DQO Process is flexible enough to be applied and adapted to a wide range of problems. The three documents chosen for this appendix are:

- *Guidance for the Data Quality Objectives Process* (EPA, 1994). Office of Research and Development, September 1994, EPA/600/R-96/055. (This document is referred to as EPA DQO.)

- *Standard Practice for Generation of Environmental Data Related to Waste Management Activities: Development of Data Quality Objectives*, D5792–95, American Society of Testing and Materials (ASTM), January 1996. (This document is referred to as ASTM DQO.)

- Module 7, "Streamlined Approach for Environmental Restoration (SAFER)" in *Remedial Investigation/Feasibility Study (RI/FS) Process, Elements and Technical Guidance*, Department of Energy (DOE), EH 94007658, December 1993. (This document is referred to as DOE SAFER.)

These comparative statements are based on an assessment of whether a particular issue was specifically and extensively addressed in the document itself. The absence of a particular issue in a document reflects the needs of its particular audience and should not necessarily be regarded as a potential deficit.

EPA DQO presents the original development of the DQO Process. Although strongly modeled on the EPA DQO approach, ASTM DQO depicts efforts by a standards organization to recast the DQO Process in a standards environment, where many opposing views should be reconciled in the production of the standard. DOE SAFER combines the DQO Process with the Observational Approach (OA) and the result is the Streamlined Approach for Environmental Restoration (SAFER). The basis of the OA is the observational method, a technique originally developed to manage uncertainty in the design and construction of subsurface facilities such as tunnels. The essence of OA is that remedial action can and should be initiated without "full" characterization of the nature and extent of the contamination.

The comparison of these documents has been divided into three specific discussion areas. First, is a general comparison of the prevalent strategy employed by each document, to assist in understanding why the separate documents might present the DQO Process in different ways. Second, a comparison of the EPA DQO Process to the DOE SAFER methodology is presented. This comparison describes the most substantive differences between documents; SAFER tends to rely more on the Observational Approach, including only some elements of the DQO Process. Finally, there is a discussion of the differences among the three documents in their presentation of decision rules and decision quality measures, which are key outputs of the DQO Process.

A.1 GENERAL COMPARISON

The approach taken by each of the three reviewed documents is best expressed by quotes from the documents themselves:

- **EPA DQO.** "The U.S. Environmental Protection Agency (EPA) has developed the Data Quality Objectives (DQO) Process as an important tool for project managers and planners to determine the type, quantity, and quality of data needed to support Agency decisions."

- **ASTM DQO.** "The DQO Process is a logical sequence of seven steps that leads to decisions with a known level of uncertainty. It is a planning tool used to determine the type, quantity, and adequacy of data needed to support a decision. It allows the users to collect proper, sufficient, and appropriate information for the intended decision."

- **DOE SAFER.** "The U.S. Department of Energy (DOE) developed the Streamlined Approach for Environmental Restoration (SAFER) as a methodology tailored to the challenges of conducting environmental restoration efforts under conditions of significant uncertainty. SAFER was developed primarily by integrating the Data Quality Objectives (DQO) Process with the Observational Approach (OA)."

In all instances, the guidance documents emphasize that the approaches are planning tools. The documents intend for users to put into up-front planning a significant effort and appropriate amount of funds to reduce subsequent costs by focusing the data collection and decision making on only those things absolutely and clearly needed to solve the problem at hand.

These three documents are based on the need to make decisions under uncertain conditions in environmental management and restoration scenarios. EPA DQO and ASTM DQO present the DQO Process in a similar "seven-step" format. For the most part, the ASTM DQO models itself after the EPA DQO document. DOE SAFER, on the other hand, does not use the "seven-step" format explicitly, but implicitly incorporates the process in describing the steps in Remedial Investigation/Feasibility Study (RI/FS) planning and through to the Remedial Design/Remedial Action (RD/RA) phase of environmental restoration.

Table A-1 compares the three approaches by different subject categories. A review of the first two subject categories, "Use" and "Audience," indicates why the presentations of the DQO Process vary from one document to the next as each organization sought to develop a presentation useful for its own needs. An examination of the next three categories ("When Applied," "Focus," and "Planning Emphasis") shows that all three documents view the DQO Process as an integral part of project planning for a data collection activity. "Explicit Stakeholder Participation" is addressed to some degree in all three documents but receives by far the greatest emphasis in DOE SAFER. The next three subject categories ("Action Oriented," "Uncertainty Addressed Directly," and "Conceptual Model") are related and will be discussed further later in this appendix. Finally, it should be noted that all three documents consider the DQO Process to be an iterative endeavor, in which applicable steps are revisited as new information is gathered during the project.

A.2 COMPARISON OF EPA DQO TO DOE SAFER

Because ASTM DQO and EPA DQO both describe the DQO Process using very similar approaches, their comparison consists primarily of describing differences in terminology and how an industry standards group applies the seven DQO steps as opposed to how a federal agency applies them. DOE SAFER represents a difference in philosophy from EPA DQO and ASTM DQO; therefore, a comparison of SAFER with EPA DQO is more substantive.

A.2.1 Safer and the Observational Approach

DOE developed SAFER to address the need, from a scientific and engineering perspective, to make decisions under uncertain conditions while maintaining progress throughout the environmental restoration process. SAFER is a methodology that is used to help streamline the RI/FS process and to manage changes to the selected remedy. SAFER is a combination of the DQO Process and the Observational Approach.

Table A-1. Key Categories Addressed in DQO Process Documents

Subject	EPA DQO	ASTM DQO	DOE SAFER
Use	EPA environmental decisions	Waste management environmental data collection	DOE Environmental Restoration (ER) Projects (CERCLA/RCRA activities)
Audience	EPA Project Managers Stakeholders	Project managers Decision makers	DOE ER Project Managers Stakeholders
When Applied	Part of the data collection planning process	Part of the data collection planning process	Part of the data collection planning process
Focus	Is the planning process	Is the planning process	Applied as an adjunct to the RI/FS, RD/RA planning processes
Planning Emphasis	Very strong	Very strong	Strong
Explicit Stakeholder Participation	Environmental community	Limited	Explicit, integral, frequent, significant
Action Oriented	Action related, but not streamlined	Action related, but not streamlined	Intended to initiate action more quickly
Uncertainty Addressed Directly	Sampling error, measurement error	Sampling error, measurement error	Sampling error, Measurement error, Probable conditions
Conceptual Model	Critical	Needed	Extremely critical
Iterative	Yes	Yes	Yes

The Observational Approach is based on the Observational Method. Originally introduced by the French soil scientist, Karl Terzaghi, the approach is described as follows (Terzaghi, 1961):

> Soil engineering projects, such as dams, tunnels, and foundations, require a vast amount of effort and labor securing only roughly approximate values for the physical constants that appear in the (design) equations. The results of the computations are not more than working hypotheses, subject to confirmation or modification during construction. In the past, only two methods have been used for coping with the inevitable uncertainties: either adopt an excessively conservative factor of safety, or make assumptions in accordance with general, average experience. The first method is wasteful; the second is dangerous.
>
> A third method is provided that uses the experimental method. The elements of this method are “learn-as-you-go”: Base the design on whatever information can be secured. Make a detailed inventory of all the possible

> differences between reality and the assumptions. Then compute, on the basis of the original assumptions, various quantities that can be measured in the field. On the basis of the results of such measurements, gradually close the gaps in knowledge, and if necessary modify the design during construction.

Used by Terzaghi, the observational procedure led to significant successes in reducing project duration and cost.

When DOE developed SAFER, the agency wanted an approach that was proactive, yet compatible and compliant with existing environmental regulations. Figure A-1 shows the SAFER process in a regulatory framework. The DOE SAFER guidance emphasizes the role of SAFER throughout the entire remediation process—from scoping and RI/FS to RD/RA. SAFER helps focus the RI/FS by emphasizing planning, making appropriate use of available data, and quickly converging on realistic remedial alternatives. SAFER streamlines the RD/RA by providing for modifying the remedy—according to preestablished contingency plans—as new information is gained.

A.2.2 Comparing SAFER to the DQO Process

SAFER targets the full sequence of decisions, from initial characterization to confirmation of the cleanup, and provides templates, checklists, and detailed definitions to help users work through SAFER elements. EPA DQO does not discuss a particular framework in which the process is applied; rather, the document states only that the process should be applied whenever environmental data collection efforts are to be undertaken. However, EPA DQO addresses the sequential nature of decision making by emphasizing that the DQO Process may be applied to many different problems throughout the investigation characterization, all the way to confirmation of the cleanup.

Although these distinctions do not result in different descriptions of the DQO Process in the two documents, they do explain why DOE SAFER places emphasis on contingency plans and monitoring plans in addition to data collection planning, while EPA DQO discusses data collection planning focused on decision making. Contingency plans are part of managing uncertainty during the actual cleanup phase. Monitoring plans are defined during the Feasibility Study phase to ensure that deviations can be detected and that the appropriate contingency plan is identified. This procedure streamlines the RI/FS by reducing the need to continually refine probable conditions by collecting data until minimal uncertainty exists. The net result is to ensure that the cost of the data collection that occurs is minimized by balancing the need to reduce uncertainty with the ability to manage it (i.e., have contingency plans ready and available). The full-spectrum approach of DOE SAFER requires consideration of what happens in the field, how the equipment performs, and what all the unknown events are that could keep the project from a clean closure.

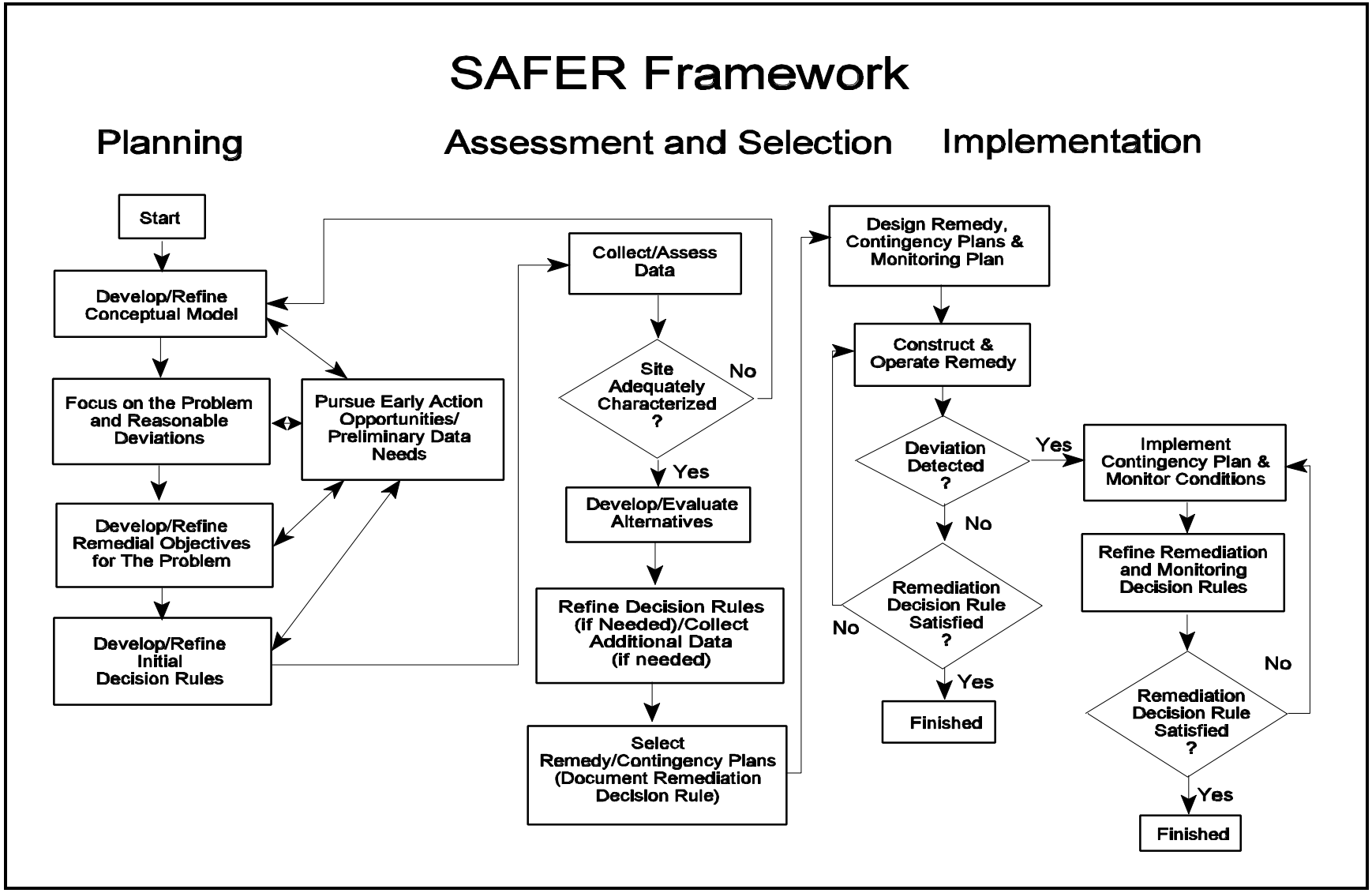


Figure A-1. DOE SAFER Framework Diagram

There are certain elements in the DQO Process and SAFER that play such an important role in making cleanup actions more streamlined and efficient that they deserve detailed scrutiny. Five elements are examined here: (1) measurement systems and sources of uncertainty, (2) optimization and trade-offs, (3) probable conditions and probable performance, (4) decision rules, and (5) reasonable deviations.

A.2.2.1 Measurement Systems and Sources of Uncertainty

DOE SAFER groups the sources of environmental restoration uncertainty into three broad areas: site conditions, remedial technology performance, and regulatory requirements. Each of these areas are defined by a set of probable conditions that researchers would say is their "best guess" at what they will find. A fourth area, measurement system limitations, is also mentioned but is addressed extensively in other parts of SAFER.

It is assumed that uncertainty in the first two areas can be *reduced* through enhancement of the measurement system.[3] DOE SAFER defines "measurement system" in general and broad terms. The measurement system encompasses "what data are to be collected, how they should be collected in the field and packaged and transported, how samples should be analyzed in the laboratory, and how data will be evaluated." This broad definition partially overlaps SAFER's definition of the role of the decision rule as "establishing . . . the types and quality of data to be collected." To manage the uncertainty that remains after the measurement system has been enhanced, contingency plans are developed. Contingency plans are designed to handle the reasonable deviations; unreasonable deviations result in an identified data gap that will need to be filled through data collection efforts. During implementation of the selected remedy, monitoring plans will identify that a deviation is occurring and say which contingency plan is necessary. Figure A-2 shows the SAFER components and the relationships just described as they relate to uncertainty.

EPA DQO distinguishes among the different components that make up SAFER's "measurement system" and provides more details than SAFER on how input from stakeholders can be used to guide the selection of the type and quality of data to collect. EPA DQO discusses the individual components— the decision rule and the decision error tolerances that should be supplied by the decision maker. EPA DQO continues in a standard statistical hypothesis testing framework to talk about the additional components of the total problem such as the specific sampling design that should be selected, the statistical tests that will be executed using the data, and the selection of the optimal sample size. Selection among the alternatives for each of these components depends on assumptions about the site, the statistical model assumed, the expected performance of the chosen statistical test (i.e., power of the test), and above all, the desired levels for the Type I and II (false rejection and false acceptance) decision errors. The desired quality or precision of the data will be dictated by how far away the decision makers expect the site data to

[7]This assumption is somewhat misleading because one could thoroughly sample a site with the best sampling equipment, and uncertainty about the performance of the technology could still exist.

be from the "cut point" or action level specified in the decision rule, along with the magnitude of allowable decision error and the indifference region (gray region) for the decision.

EPA DQO has a more limited use of the term measurement error. The guidance states that when one cannot know the true value of a population parameter, it is because there is sampling error (natural variability in true state of the environment) and "measurement error" (which is a combination of errors that occur in the collection, handling, preparation, analysis, and reduction of sample data). The combination of sampling error and measurement error is called total study error.

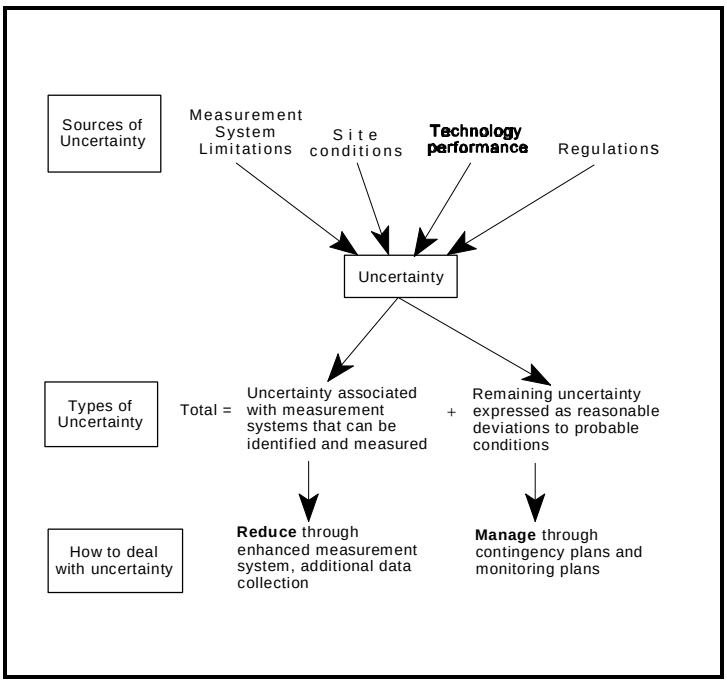


Figure A-2. Components of Uncertainty in DOE SAFER

A2.2.2 Optimization/Tradeoffs

DOE SAFER does not provide detailed guidance for making the tradeoff between reducing uncertainty (improving the measurement systems) and managing uncertainty (increasing the reliance on improved contingency planning). SAFER does not explicitly use EPA DQO's statistical hypothesis testing framework for discussing decision errors; rather, SAFER's only guidance is that stakeholders should "mutually agree" on trade-offs. SAFER does state that contingency plans need to be developed for only the reasonable deviations, as judged by the stakeholders. This places a limit on how much total uncertainty can be considered.

Similarly, EPA DQO provides only limited guidance on what to do if the sample size and/or data quality requirements cannot be met within the stakeholders' budget and time constraints and their limits on decision errors. EPA DQO addresses such balancing of tradeoffs in DQO Step 7, "Optimize the Design." EPA DQO includes in the optimization not only need to assess the tradeoffs for balancing increased sampling costs with reduced decision errors, but also the need to optimally match the problem and decision statements to the sampling designs, statistical tests, and sample size calculations.

EPA DQO does not explicitly consider SAFER's last two sources of uncertainty (i.e., technology performance and regulatory requirements), so the optimization that is discussed in EPA DQO is the optimization of the sampling design, which refers to the tradeoff of cost versus

decision-error reduction. SAFER expands on the full range of tradeoffs required, from planning to remedial action.

A2.2.3 Probable Conditions/Probable Performance

SAFER and EPA DQO differ as to how much prior information and confidence in the conceptual model stakeholders should bring to the process. SAFER alludes to some statistical decision theory tools that depend more heavily on the quality of prior information provided by stakeholders; however, such tools are not explicitly used in the statistical hypothesis testing framework around which EPA DQO is built.

SAFER requires the stakeholders to provide estimates of site conditions, to the point of making estimates of what contaminant concentrations may exist. This is required because they should then specify what are the reasonable deviations (versus unreasonable deviations) that can be expected at the site. Stakeholders should also specify the probable performance of the remedial alternatives. EPA DQO uses a conceptual model as the source of the estimate of expected total study error. However, specific levels of expected contamination are not required input as they are in SAFER. SAFER uses the difference between the estimated and measured responses to determine whether the environmental system is best represented by the probable condition or by a deviation—which would trigger the contingency plan. The ability to determine significant differences is critical to the success of SAFER.

A.2.2.4 Decision Rules

DOE SAFER uses the term “decision rules” in a much more general context than does EPA DQO. For EPA DQO, a decision rule is a structured statement of the following form: “IF (the true population parameter of interest) is (greater than/less than/equal to) the action level, THEN take action #1.” DOE SAFER gives a much more general description of decision rules, saying they “summarize how uncertainty will be reduced by the data measurement system”; “are formulated to clearly identify data needs and data uses”; and “are used to identify data that are collected during monitoring to identify deviations and to determine when the remedial goals have been accomplished.”

One may interpret the relationship between DOE SAFER and EPA DQO such that many of the functions ascribed to the decision rule in SAFER are in fact implemented by the DQO Process. For example, whereas SAFER says that the decision rules summarize how uncertainty in the estimates of population parameters based on sample data can be reduced, EPA DQO states that this uncertainty can be reduced by: (1) choosing an appropriate sampling design selected on the basis of what problem is being addressed and how the conceptual model hypothesizes the contamination is spatially distributed, (2) choosing appropriate sample collection and analysis equipment, and (3) choosing larger sample sizes. In addition, the decision maker specifies what decision error is “acceptable” and specifies the size of the gray region—both of which will be taken into consideration in selecting an optimal design and determining the operational form of

the decision rule. In this interpretation, the EPA DQO Process can be seen as being embedded as a tool or functional routine within the larger SAFER process.

A.2.2.5 Reasonable Deviations

Identifying which deviations are reasonable and preparing contingency plans to address them are primary SAFER techniques. This process of identifying reasonable deviations streamlines the RI/FS by lessening the need to attempt to eliminate uncertainty. If contingency plans are prearranged in the event a reasonable deviation from probable conditions is encountered, action can continue in the field under a wide variety of conditions. SAFER thus claims to have a "bias for action."

Somewhat comparable concepts to SAFER's reasonable deviations are boundary conditions and the "gray region" in the DQO Process. However, the consequences of being outside the limits are very different. When EPA DQO decision makers specify a gray area, they are saying they are not concerned about or are not willing to spend the resources to control decision errors in the gray area. Outside the gray region, the decision makers specify acceptable limits on decision errors. When SAFER stakeholders specify the reasonable deviation interval, they are saying site conditions or performance indicators inside this interval can be addressed through contingency plans. Values outside this interval are unreasonable deviations and are unlikely (or will not affect remedial activities) and are not amenable to contingency planning.

A.3 DECISION RULES AND DECISION QUALITY MEASURES

The initial steps of the DQO Process are designed to focus the investigation on what the real problem is, what decisions need to be made to solve the problem, and what the boundaries of the decision are. After this has been accomplished, it is possible to develop an ideal decision rule and specify measures of desired quality for an operational decision rule considering the reality of uncertainty. These two activities are presented somewhat differently in each of the documents (EPA DQO Chapters 5-7, ASTM DQO Section 6.6-6.8, and DOE SAFER Submodule 7.2) and are a potential source of confusion among those comparing the documents.

A.3.1 Develop an Ideal Decision Rule

Readers should be careful to keep in mind the difference between (1) the concept of the *ideal* decision rule and *desired* decision quality, and (2) the concept of the *operational* decision rule and *achievable* decision quality. There is a natural tendency to confuse the two concepts as the three documents use different terminologies for similar concepts. An *ideal* decision rule does not consider uncertainty but clearly states the kind of decision that the planning team desires to make. An *operational* decision rule will actually be applied to the data and will take into account the uncertainty that will enter into the decision process. The following paragraphs explain how these concepts are treated differently among EPA DQO, ASTM DQO, and DOE SAFER.

EPA DQO: EPA DQO is very explicit about defining the *ideal* decision rule (and only the ideal decision rule) in Step 5. It defines the *ideal* decision rule as:

If (the parameter of interest) is greater than the (action level),
then (take appropriate action for the problem),
otherwise (take appropriate action for no problem).

ASTM DQO: This document addresses an *operational* decision rule instead of an *ideal* decision rule. This *operational* decision rule depends on the concept of acceptable decision error tolerances, an idea the document has not introduced at this stage. This point also highlights an instance of two documents using the same terminology (and identical glossary definitions) for two very different concepts. ASTM DQO uses action level in the sense of the "to-be-determined decision point" in the *operational* rule, whereas EPA DQO uses action level in the sense of "level of concern" in the *ideal* decision rule. This difference is a critical distinction and causes comprehension problems for those comparing both documents.

DOE SAFER: This module does not explicitly discuss either form—*ideal* or *operational*—of the decision rule. Instead, one page is dedicated to a discussion of the benefits of having decision rules. However, it is unclear if DOE SAFER is presenting *ideal* or *operational* decision rules.

A.3.2 Measures of Desired Quality

Once the statement of the *ideal* decision rule has been completed, it becomes the responsibility of all stakeholders to agree on (or the decision maker to specify) some measure of the desired quality of an *operational* decision rule that takes uncertainty into account. All three documents generally follow the same underlying ideas in presenting measures of desired decision quality. However, there are both conspicuous and subtle differences in the presentations that may serve to confuse those new to the concept. In addition, understanding may be difficult because of identical terminology with different meanings from one document to the next.

There are several fundamental quality measures which can be displayed on a basic decision quality measure graph (see Figure A-3) and which are discussed below. Using this basic graph as a model, it is possible to discuss the derivation of the decision quality measures for each of the three documents using approximately the same scale to make cross-comparisons apparent. Figure A-4 shows an example of the quality measure depictions for each document and Table A-2 contains a summary of this discussion.

- *Horizontal axis*: The axes have approximately the same function in each instance; however, each document gives the axis a somewhat different label.

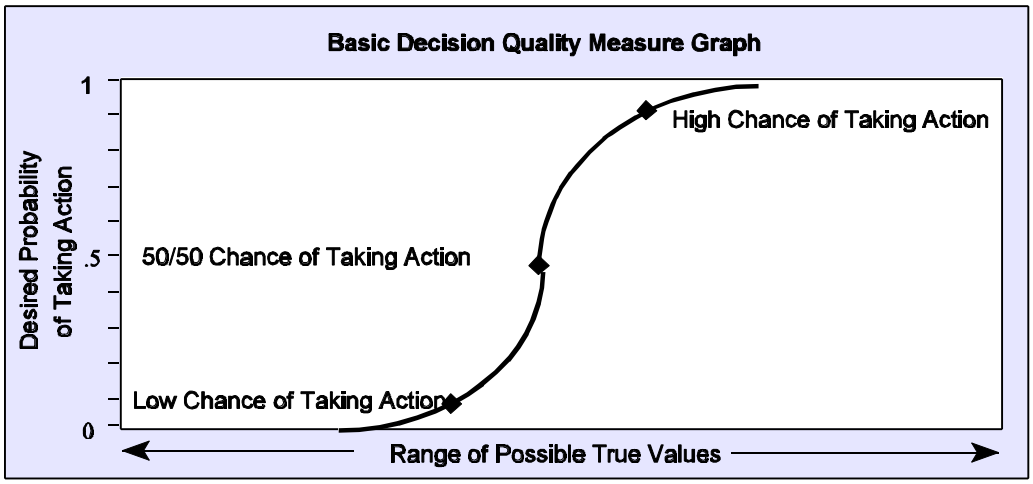


Figure A-3. Basic Decision Quality Measure Graph

- *Level of Concern*: EPA DQO has labeled the level of concern "Action Level" and states that it is one of the boundaries of the "Gray Region." ASTM DQO calls it "Regulatory Threshold." It is somewhat unclear if DOE SAFER use of "cut point" is the same as "level of concern," but the way it is presented graphically argues otherwise. (See discussion of vertical axis and probability curve below.)

- *Vertical axis*: This element serves the same function in both EPA DQO and ASTM DQO although each document gives it a somewhat different label. Again, it is unclear how DOE SAFER uses this axis. It appears that in its presentation it has simply "folded" the basic graph at the 50/50 point and is using qualitative labeling of the probabilities. (See discussion of probability curve.)

- *Probability Curve*: For each possible true value, the probability curve shows the desired probability of concluding there is a problem and that some action should be taken. This element is present in all three documents. ASTM DQO presents it as a continuous function, DOE SAFER has folded the probability curve at the 50/50 point, and EPA DQO has chosen to present the curve in discrete portions (i.e, only at several points). The sections of the probability curve between the action level and another selected point are essentially hidden behind the gray region, that subset of the true values of the parameter where relatively large decision error rates are considered tolerable. The left and right tails of the probability curve have simply been turned into "step functions" at selected points, one on each edge of the gray region, and one farther out on each tail.

- *"Low" and "High" Probability Points*: EPA DQO and ASTM DQO use low and high probability points (i.e., a true value where it is important to have a low probability of taking action and a true value where it is important to have a high probability of taking action) to bound their gray region and for use in statistical

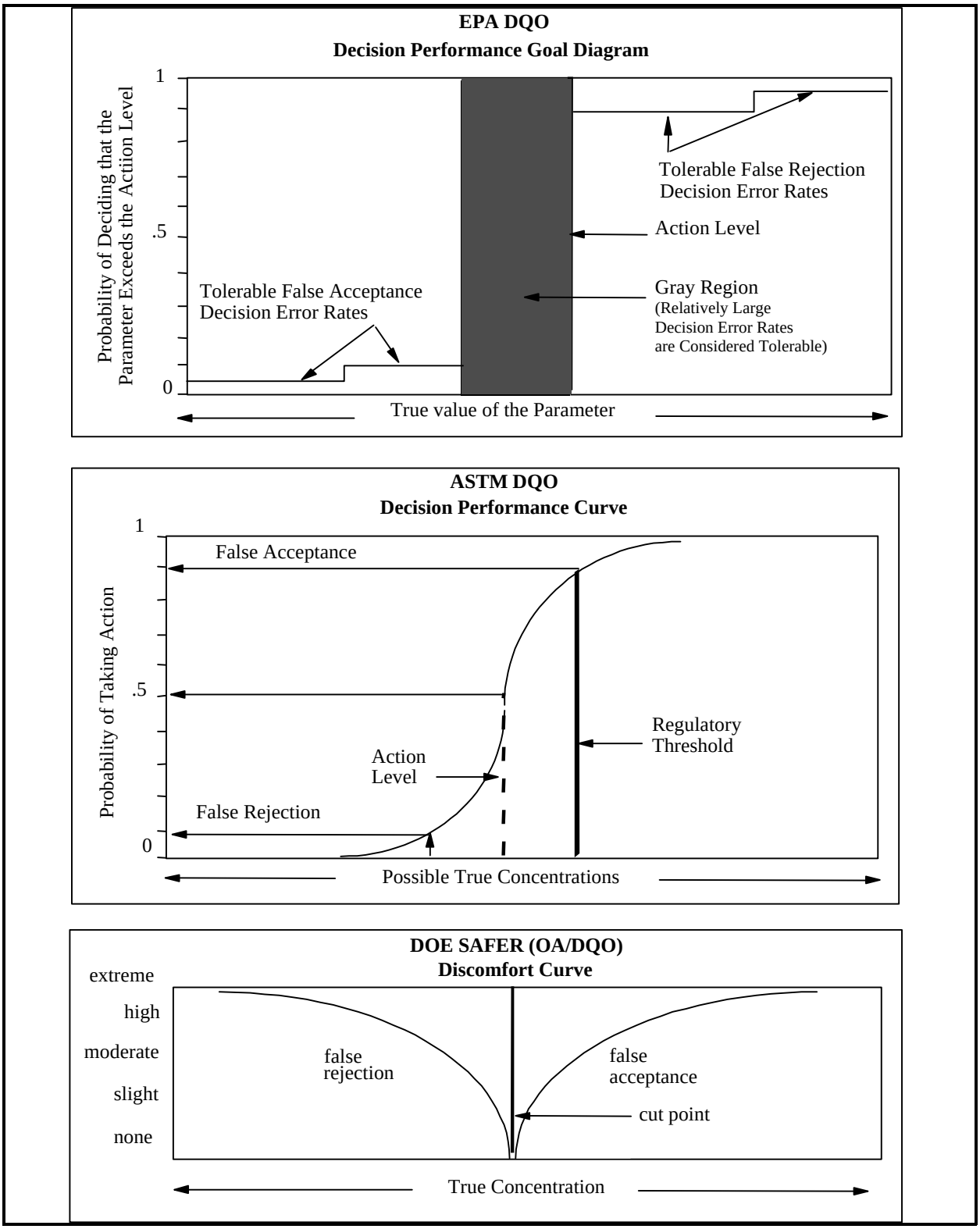


Figure A-4. Decision Quality Measure Graphs for EPA DQO, ASTM DQO, and DOE SAFER

Table A-2. Fundamental Elements of the Basic Decision Quality Measures

	EPA DQO	ASTM DQO	DOE SAFER
Horizontal Axis	True Value of the Parameter	Possible True Concentration	True Concentration
Level of Concern	Action Level	Regulatory Threshold	Cut Point
Vertical Axis	Probability of Deciding That the Parameter Exceeds the Action Level	Probability of Taking Action	Unlabeled, Folded, Qualitative
Probability Curve	Discrete, Gray Region	Straightforward	Folded
Probability Points	Statistical Hypothesis, Gray Region Boundaries	Statistical Hypothesis	(Not explicitly used)
50/50 Probability Point	Covered by Gray Region	To-Be-Determined Decision Point	Cut Point, "Folding" Point
False rejections and False acceptances	Tied to statistical null hypothesis	Tied to taking action	Unclear

hypothesis testing. One or the other of these will serve as the null hypothesis in the statistical framework, depending on the importance of taking action at the action level/regulatory threshold. (Note: The graphical presentation in this document for EPA DQO is based on a null hypothesis (baseline assumption) that the action level/regulatory threshold has been exceeded. EPA DQO also covers the possibility of the opposite null hypothesis—that the action level is not being exceeded—whereas ASTM DQO does not.) DOE SAFER does not explicitly make use of these points.

- *50/50 Probability Point*: The 50/50 probability point is not used explicitly by EPA DQO since it is covered up by the gray region. DOE SAFER uses the 50/50 point to fold the basic graph and labels this point the cut point, i.e., the true value where either decision is considered acceptable. The 50/50 probability point ("Action Level") used in ASTM DQO represents the "to be determined" decision point. A resulting characteristic of this operational rule decision point is that it will have a 50/50 probability for the true value that it happens to coincide with.

A final consideration is that the use of the qualifiers “false positive” (false rejection error) and “false negative” (false acceptance error) are not identical among the three documents, which may lead to some confusion. EPA DQO uses both qualifiers in the context of statistical hypothesis testing. With this approach a false rejection decision error is made when the null hypothesis is incorrectly rejected and a false acceptance decision error is made when the null hypothesis is incorrectly accepted. The ASTM DQO document takes a different approach, tying the definition of these two terms to whether or not action should have been taken. With this approach a false rejection decision error is made when action is taken that was unnecessary and a false acceptance decision error is made when no action is taken although it should have been. Depending on how the null hypothesis is stated, these different sets of definitions could mean the same thing or they could mean exactly the opposite of each other. It is unclear how DOE SAFER uses these terms, although the text suggests that it may be using the EPA DQO approach.

A.4 CONCLUSION

With regard to decision rules and decision quality measures, the user of any one of these documents should be able to deal with general concepts, as all three documents use somewhat similar underlying ideas. However, any user who tries to reconcile the differences between any two documents or gets into a discussion with a user of another document about a particular detail will need to be wary of the differences in presentation, terminology, and specific methodology that have been described in this section, so as to avoid miscommunication.

The EPA’s Quality System requires the use of a systematic planning process but does not mandate the use of the EPA DQO Process. It does, however, highly recommend the adoption of the DQO Process and use of the DQO Process fully meets the Agency’s requirements. Use of other planning processes are acceptable, but care should be taken to ensure misunderstanding of key techniques does not occur.

APPENDIX B

GLOSSARY OF TERMS USED IN THIS DOCUMENT

action level: the numerical value that causes a decision maker to choose one of the alternative actions (e.g., compliance or noncompliance). It may be a regulatory threshold standard, such as a maximum contaminant level for drinking water; a risk-based concentration level; a technological limitation; or a reference-based standard. Note that the action level defined here is specified during the planning phase of a data collection activity; it is not calculated from the sampling data.

alternative condition: a tentative assumption to be proven either true or false. When hypothesis testing is applied to site assessment decisions, the data are used to choose between a presumed baseline condition of the environment and an alternative condition. The alternative condition is accepted only when there is overwhelming proof that the baseline condition is false. This is often called the alternative hypothesis in statistical tests.

baseline condition: a tentative assumption to be proven either true or false. When hypothesis testing is applied to site assessment decisions, the data are used to choose between a presumed baseline condition of the environment and an alternative condition. The baseline condition is retained until overwhelming evidence indicates that the baseline condition is false. This is often called the null hypothesis in statistical tests.

bias: the systematic or persistent distortion of a measurement process that causes errors in one direction (i.e., the expected sample measurement is different from the sample's true value.

boundaries: the spatial and temporal conditions and practical constraints under which environmental data are collected. Boundaries specify the area or volume (spatial boundary) and the time period (temporal boundary) to which a decision will apply. Samples are collected within these boundaries.

data collection design: see sampling design.

data quality objectives (DQOs): qualitative and quantitative statements derived from the DQO Process that clarify study objectives, define the appropriate type of data, and specify tolerable levels of potential decision errors that will be used as the basis for establishing the quality and quantity of data needed to support decisions.

data quality objectives process: a quality management tool to facilitate the planning of environmental data collection activities. Data quality objectives are the qualitative and quantitative outputs from the DQO process.

decision error: the error that occurs when the data mislead the site manager into choosing the wrong response action, in the sense that a different response action would have been chosen if the site manager had access to unlimited "perfect data" or absolute truth. In statistical tests, decision errors are labeled as false rejection or false acceptance depending on the concerns of the decision maker and the baseline condition chosen.

defensible: the ability to withstand any reasonable challenge related to the veracity or integrity of project and laboratory documents and derived data.

false rejection decision error: the error that occurs when a decision maker rejects the baseline condition (null hypothesis) when it actually is true. Statisticians usually refer to the limit on the possibility of a false rejection error as alpha (α), the level of significance, or the size of the critical region, and it is expressed numerically as a probability.

false acceptance decision error: the error that occurs when a decision maker accepts the baseline condition when it is actually false. Statisticians usually refer to the limit on the possibility of a false acceptance decision error as beta (β) and it is related to the power of the statistical test used in decision making.

gray region: the range of possible parameter values near the action level where the cost of determining that the alternative condition is true outweighs the expected consequences of a decision error. It 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. It is sometimes referred to as the region where it is "too close to call."

judgmental sampling: a subjective selection of sampling locations based on experience and knowledge of the site by an expert without the use of a probabilistic method for sample selection.

limits on decision errors: the acceptable decision error rates established by a decision maker. Economic, health, ecological, political, and social consequences should be considered when setting limits on decision errors.

mean: a measure of central tendency. A population mean is the expected value ("average" value) from a population. A sample mean is the sum of all the values of a set of measurements divided by the number of values in the set.

measurement error: the difference between the true or actual state and that which is reported from measurements.

median: a measure of central tendency, it is also the 50th percentile. The sample median is the middle value for an ordered set of *n* values; represented by the central value when *n* is odd or by the average of the two most central values when *n* is even.

medium: a substance (e.g., air, water, soil) that serves as a carrier of the analytes of interest.

natural variability: the variability that is inherent or natural to the media, objects, or people being studied.

parameter: a descriptive measure of a characteristic of a population. For example, the mean of a population (μ).

percentile: a value on a scale of 100 that indicates the percentage of a distribution that is equal to or below it. For example, if 10 ppm is the 25th percentile of a sample, then 25 percent of the data are less than or equal to 10 ppm and 75 percent of the data are greater than 10 ppm.

performance curve: the probability of deciding that the parameter of interest is greater than the action level over the range of possible population parameters. It is similar in concept to a statistical power curve. The performance curve is used to assess the goodness of a test or to compare two competing tests.

planning team: the group of people who perform the DQO Process. Members include the decision maker (senior manager) site manager, representatives of other data users, senior program and technical staff, someone with statistical expertise, and a QA and QC advisor (such as a QA manager).

population: the total collection of objects or people to be studied and from which a sample is to be drawn.

power curve: the probability of rejecting the baseline condition over the range of the population.

probabilistic sampling: a random selection of sampling locations that allows the sampling results to be extrapolated to an entire site (or portion of the site).

quality assurance (QA): an integrated system of management activities involving planning, implementation, documentation, assessment, reporting, and quality improvement to ensure that a product, item, or service is of the type and quality needed and expected by the customer.

QA Project Plan: a document describing in comprehensive detail the necessary QA, QC, and other technical activities that should be implemented to ensure that the results of the work performed will satisfy the stated performance criteria.

quality control (QC): the overall system of technical activities that measures the attributes and performance of a process, item, or service against defined standards to verify that they meet the stated requirements established by the customer; operational techniques and activities that are used to fulfill requirements for quality.

range: the numerical difference between the minimum and maximum of a set of values.

sample: (i) a single item or specimen from a larger whole or group, such as any single sample of any medium (e.g., air, water, soil); or

(ii) a group of samples from a statistical population whose properties are studied to gain information about the whole.

The definition is decided by context of usage.

sample variance: a measure of the dispersion of a set of values. Small variance indicating a compact set of values; larger variance indicates a set of values that is far more spread out and variable.

sampling: the process of obtaining a subset of measurements from a population.

sampling design: a design that specifies the final configuration of the environmental monitoring effort to satisfy the DQOs. It includes what types of samples or monitoring information should be collected; where, when, and under what conditions they should be collected; what variables are to be measured; and what QA and QC components will ensure acceptable sampling error and measurement error to meet the decision error rates specified in the DQOs. The sampling design is the principal part of the QA Project Plan.

sampling design error: the error due to observing only a limited number of the total possible values that make up the population being studied. Sampling errors are distinct from those due to imperfect selection; bias in response; and mistakes in observation, measurement, or recording.

statistic: a function of the sample measurements (e.g., the sample mean or sample variance).

total study error: the sum of all the errors incurred during the process of sample design through data reporting. This is usually conceived as a sum of individual variances at different stages of sample collection and analysis.

variance: see sample variance.

APPENDIX C

JUDGMENTAL SAMPLING
DQO CASE STUDY: ACCONADA STORAGE FACILITY

Background

This case concerns a hypothetical commercial storage facility operated by Acconada, Inc. and located in Northern Florida. The surrounding area is a mix of light industrial, commercial, and residential properties. Acconada, Inc. owns approximately 25 acres on which hazardous and non-hazardous materials have been stored, handled, and sold. Building A was the hazardous materials and hazardous waste storage warehouse. This case study addresses only the early stages of site assessment for Building A and the grounds immediately surrounding Building A.

Acconada, Inc. has been in operation since the 1940s, when the storage area was built for general warehousing operations. Documentation indicates that operations have involved the receipt of hazardous materials since at least the early 1980s, as well as the receipt of non-hazardous materials. During a recent reconnaissance visit to the site, a group of 55-gallon drums were observed in the unpaved area immediately adjacent to the entrance to Building A. The soil around the drums was stained, indicating possible leaking. The interior of Building A is currently undergoing closure under the Resource Conservation and Recovery Act (RCRA) and is therefore not included in this case study. Figure C-1 depicts Building A, the immediate surroundings, and

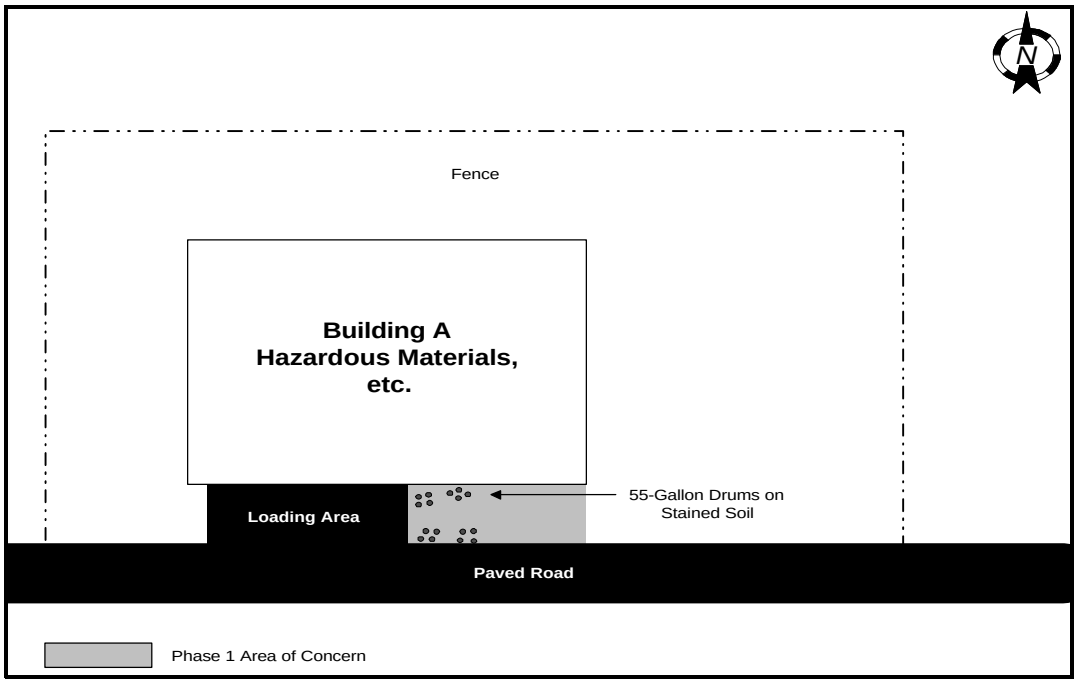


Figure C-1. Building A and Surrounding Area - Phase 1

the location of the leaking drums. This figure does not show the entire 25-acre Acconada site, as only Building A and its immediate surroundings are addressed in this case study.

Hazardous wastes have typically been received at the Acconada facility in 55-gallon drums and other waste containers. According to Acconada records, there have been no previous (known or suspected) releases of hazardous substances to the environment. This case study begins with an initial site assessment under CERCLA and in coordination with the ongoing RCRA activities at the site.

Step 1: State the Problem

This step summarizes the contamination problem, identifies the planning team, develops a conceptual site model, identifies exposure scenarios, and determines the resources available for the study.

Identify Members of the Planning Team — The planning team includes the EPA Regional Remedial Project Manager (RPM); a management representative from Acconada, Inc.; and technical staff including a field sampling expert, a chemist from the analytical laboratory, a risk assessor, and a geologist with expertise in sampling designs for surface soil. A town council member joined the planning team to represent the local government and the interests of nearby residents and businesses. The primary decision maker is the Acconada manager. The RPM is responsible for approving documents and plans on behalf of EPA as well as providing guidance and suggestions.

Develop the Conceptual Site Model — There is evidence that minor spills or leaks occurred at Acconada based on visible staining of the surface soil around the containers in front of Building A. Figure C-2, the Conceptual Site Model (CSM), illustrates the possible pathways and exposure routes that the planning team considered. According to the available information and statements of the workers, the leaking containers contain a pesticide, Chlordane. Chlordane, a chemical commonly found at NPL sites, was used as a pesticide from 1948 until 1988. All uses of Chlordane were banned by the EPA in 1983 except for termite control, which was banned in 1988. Chlordane is a thick liquid which ranges in color from colorless to amber. Chlordane adheres strongly to soil particles at the surface, is not likely to enter groundwater, breaks down very slowly, and can remain in soil for over 20 years. Exposure to Chlordane can affect the nervous system, digestive system, and liver in people and animals.

Define the Exposure Scenarios — The 25-acre Acconada site has been proposed for sale and transfer for residential and commercial mixed-use development. The pending sale and the potential risk to humans prompted the concern regarding whether the surface and subsurface soils pose an unacceptable risk to human health and the environment. Potential human receptor populations include construction workers, future residents, or visitors who may come in contact with the contaminated soils or airborne particulates. Access to the facility continues to be

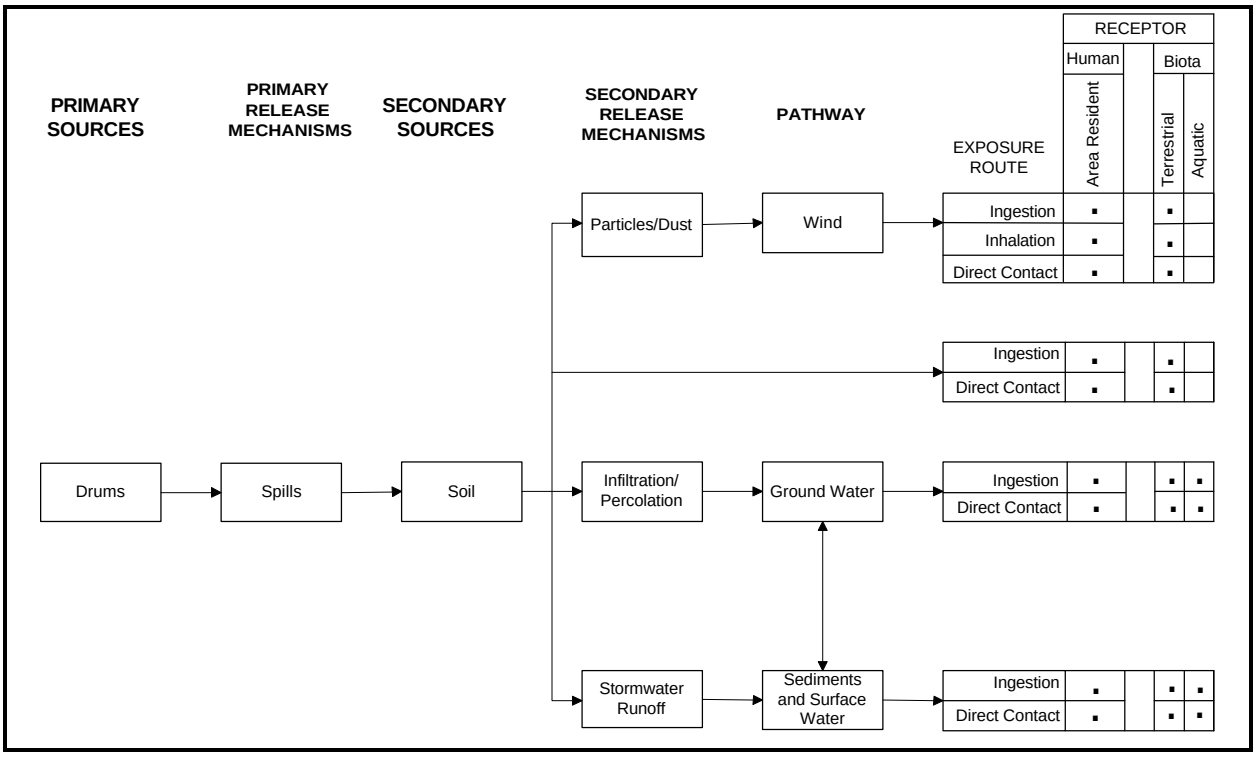


Figure C-2. **Conceptual Site Model**

restricted; therefore, exposure is limited to personnel working at the facility and individuals permitted to enter the area. Exposure may occur if these individuals have direct physical contact with hazardous substances or if they inhale airborne particles. Individuals involved in activities that involve disturbance of the drums, soil, or vegetation in contaminated areas (such as grounds keepers or backhoe operators) also may be exposed to contaminants through inhalation.

Although Chlordane is unlikely to enter groundwater, the team confirmed that groundwater issues are addressed in a separate and more complex study that involves other buildings on the site with multiple contaminants of concern. This more complex study was undertaken to address the groundwater issues as a result of other site activities not associated with Building A; and these issues are therefore, beyond the scope of this case study. However, the team considered it important to cooperate with and be informed of other site investigation activities to avoid duplication of effort and inconsistencies.

Specify Available Resources and Constraints — Although a specific budget was not initially established, the planning team members all expressed their interest in an efficient and acceptable sampling and analysis approach. The planning team recognized that the cost of investigating and remediating the site was an important component of the cost analysis for the overall redevelopment business decision. The planning team agreed to submit the estimated cost of this part of the overall site investigation for Acconada management approval before significant resources were committed for data collection or remediation.

Although there were not specific budgetary constraints, there was a critical time constraint. Acconada management decided to remove the leaking drums within two weeks, depending on equipment and personnel availability. Acconada management would like to remove any highly contaminated soil simultaneously with the removal of the drums so that acute risks could be mitigated quickly and cost effectively. Therefore, any sampling and decision making about the soil associated with the drums had to occur concurrently with these drum removal operations. The planning team also recognized that there would need to be a second phase of investigation to more thoroughly investigate risks.

Step 2: Identify the Decision

In this step, the principal study question will be made into a decision statement that will address the contamination problem.

Identify the Principal Study Question — For the investigation of acute risks, the team identified the principal study question as: "Does this site pose a serious and/or immediate threat to human health and the environment?"

Identify Alternative Actions that Could Result from Resolving the Principal Study Question — The possible outcomes or actions that may result include:

- remove the highly contaminated soil associated with the leaking drums, and/or implement institutional or engineering controls to restrict access and potential exposure; versus

- leave the soil in place and include the area as part of the next phase of investigation.

Combine the Principal Study Question and Alternative Actions into A Decision Statement — The team combined the alternative actions and the principal study question into a decision statement:

> "Determine whether this site poses a serious and/or immediate threat to human health or the environment and thus requires an immediate response action."

Organize Multiple Decisions — The team decided to clarify the relationships among the multiple decisions based on the information available at this time. The team considered it very likely that following the immediate removal of the drums and associated soil if necessary, further investigation would be required to determine whether or not the contamination posed an unacceptable risk to human health and the environment. Therefore, the team identified two phases of investigation (see Figure C-3). Phase 1 would address the drum removal and potential removal of soil in close proximity to the drums. Phase 2 would address the larger area surrounding Building A. The team recognized that subsequent decisions would have to be made if the Phase 2 area was found to pose an unacceptable risk. However, the team decided that those

subsequent decisions could be addressed better after evaluating the results of Phase 2; hence those subsequent decisions are not shown in Figure C-3 and are not addressed in this case study.

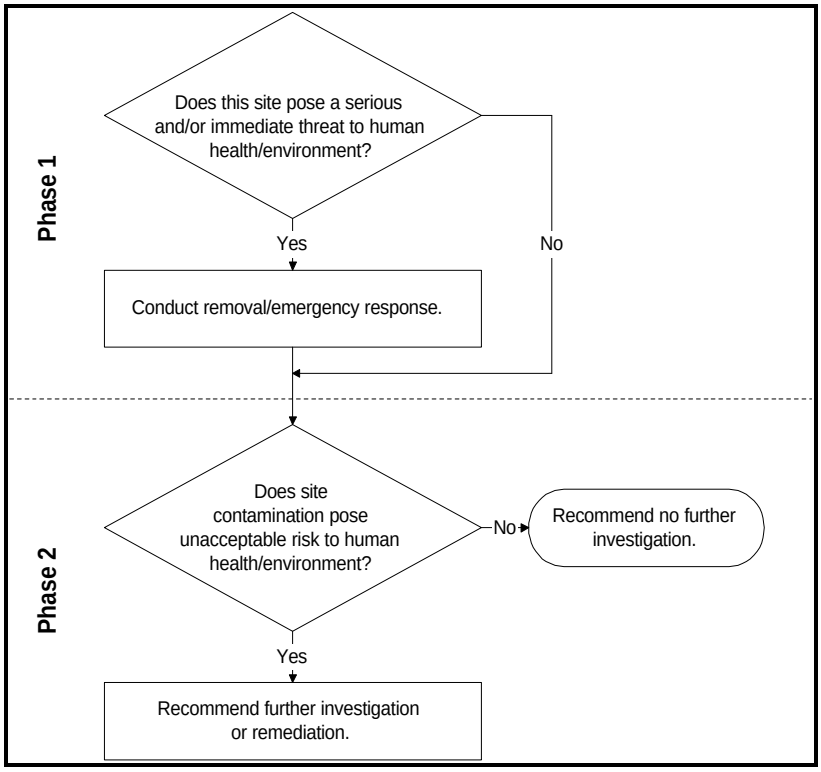


Figure C-3. Decision Diagram

Step 3: Identify the Inputs to the Decision

Here the study team assembles all the relevant information that bears on the decision statement including information on the availability of chemical methods, detection limits, and specific sources for needed information.

Identify the Information that Will Be Required to Resolve the Decision Statement — The team gathered the existing information, which included documentation of site activities, reports from the site workers, and photographs of the site. The team found the information consistent with their current understanding of the problem.

The team documented the visible soil stains, presumably from the drums, but they did not discover any existing analytical data from the site. The team decided that soil samples would be required to confirm the reported contamination. If the contamination posed a serious or immediate threat, then the team anticipated that a removal action would be needed. Specifically, the team would require data to confirm the contamination and the contaminant(s) of concern.

Determine the Sources for each Item of Information Identified —Much of the information needed was already available to the planning team as described above. The environmental data would need to be generated through sampling and analysis of the soil and contents of the drums in front of Building A.

Identify the Information Needed to Establish the Action Level—The information needed to establish the action level includes the potential chemicals of concern (Chlordane) and existing state and federal requirements or recommendations for clean-up levels.

Confirm that Appropriate Analytical Methods Exist to Provide the Necessary Data — The severe time constraint for obtaining results and making a decision prompted the team to look into field measurement methods. Consultation with field sampling and laboratory analysis experts confirmed that this investigation was an appropriate candidate for applying field measurement techniques. SW846 Draft Method 4041, Soil Screening for Chlordane by Immunoassay, is a semi-quantitative procedure for determining whether Chlordane is present above or below specific concentrations. The lower limit of detection (LLD) is 14 ppb. The lower end of the test range can be calibrated as low as 20 ppb, which is the lower limit of quantitation (LLQ); the upper end of the test range can be as high as necessary. When exact concentrations of Chlordane are required, traditional laboratory methods such as Method 8081 (gas chromatography) or Method 8270 (gas chromatography/mass spectrometry) can be used. In addition, these methods can be used to provide laboratory confirmation of the Method 4041 results.

Given the short time frame the planning team had in which to make a decision about the soil surrounding the drums, the team agreed that they would use the screening method. Prior to making this decision, the team reviewed the performance of the screening method to make sure the method would be appropriate for their intended use. In addition, the team confirmed that test kits are commercially available for Method 4041 (e.g., EnviroGard™). EnviroGard™ is accepted by EPA for SW846 Draft Method 4041. The average cost per sample is less than $20 (materials only) and approximately 16 tests can be run in less than an hour. Past field tests with this test kit indicated a high degree of reliability.

The EnviroGard™ Chlordane in Soil Test Kit uses polyclonal antibodies that bind with either Chlordane or Chlordane-Enzyme Conjugate. More specifically, a 10 gram soil sample containing Chlordane is added to a test tube containing Assay Diluent. Then, Chlordane-Enzyme Conjugate is added, which competes with Chlordane for antibody bonding sites. The test tube is incubated for 15 minutes after which time the unbound molecules are washed away. A clear solution of chromogenic substrate is added to the test tube which is converted to a blue color in the presence of bound Chlordane-Enzyme Conjugate. A sample with a low Chlordane concentration allows the antibody to bind many Chlordane-Enzyme Conjugate molecules resulting in a dark blue solution. Conversely, in a high Chlordane concentration, fewer Chlordane-Enzyme Conjugate molecules are bound to the antibodies resulting in a lighter blue solution. In other words, color development is inversely proportional to Chlordane concentration.

The Chlordane level in a sample of unknown concentration is determined by comparison to assay calibrator levels by visual comparison or with a spectrophotometer.[1] Each test kit includes three standard calibrators for Chlordane at 20, 100, and 600 ppb. However, soil extracts may be diluted to allow for interpretation at different concentrations of Chlordane. For example, if the expected range of Chlordane concentration exceeds 600 ppb, the soil extract may be diluted 1:10 in 90% methanol (as documented in the assay procedure). This would allow for interpretation at Chlordane concentrations of 200, 1000, and 6000 ppb (i.e., .2, 1, and 6 ppm).

---

[1]Strategic Diagnostics Inc. 1997. *User's Guide for EnviroGard™ Chlordane in Soil Test Kit.*

While the range of concentration can be varied according to site conditions, note that the ratio for interpreting Chlordane concentration remains unchanged. For example, samples can be interpreted by comparison to standard calibrators for 20, 100, 600 ppb or .2, 1, 6 ppm; the concentrations for each set of calibrators is distinct, yet the ratio within each set is equivalent (i.e., 1:5:30). The action level or concentration of interest should be set at the middle calibrator because the precision of an immunoassay is highest in the center of the working range.

One important limitation of the test kit is that the test kit cannot differentiate between Chlordane and other structurally similar compounds, but detects their presence to differing degrees. In other words, non-target compounds are cross-reactive in that they will compete for the finite number of antibody binding sites.[2] Cross-reactivity will impact the color development and yield erroneous results. Specifically, the presence of Endrin, Endosulfan I and II, Dieldrin, and Heptachlor will cause positive test results at lower concentrations than Chlordane alone. However, Aldrin, Toxaphene, Lindane, Alpha-BHC, and Delta-BHC require higher concentrations than Chlordane for a positive result.[3] For this reason, the test kits are appropriate only when there is existing information on the COC. Accurate information on the concentration range is less critical because the soil extract can be diluted and the assay performed again should the sample test tube contain less color than the highest calibration tube (i.e., when the Chlordane concentration exceeds that of the highest calibrator).

Step 4: Define the Boundaries of the Study

The desired outputs from this step are: a detailed description of the characteristics that define the population of interest, the spatial component of media addressed by the decision, the time period in which samples will be taken and to which decisions apply, the smallest subarea affected by the decision, and any practical constraints that could impact the sampling plan.

Specify the Characteristics that Define the Population of Interest — The Acconada site manager was interested in identifying the worst-case conditions in the area of concern. Therefore, the team agreed that the target population for this investigation would be the surface soil that was visibly stained. The team also agreed that the characteristic of interest for that target population was the concentration of Chlordane, although they recognized that they might find other contaminants. The team consulted the Superfund Soil Screening Guidance to define "stained surface soil" as the top 2 centimeters of discolored soil in the area of concern. The team recognized that the leaked Chlordane might have penetrated further down into the soil, and that any soil removal operations would have to consider greater depths of removal. However, sampling only the top 2 centimeters would minimize the chance of diluting the sample with unstained soil that might lie under the surface.

---

[2]USEPA. October, 1996. *Region I, EPA-New-England, Immunoassay Guidelines for Planning Environmental Projects*.

[3]Strategic Diagnostics Inc. 1997. *User's Guide for EnviroGard™ Chlordane in Soil Test Kit*.

Define the Spatial Boundaries of the Decision — The boundaries for the site included the area in front of Building A (excluding the paved loading dock area) where the drums are located and where the stained soil is visible (see Figure C-1).

Define the Temporal Boundaries of the Decision —The planning team did not anticipate specific seasonal or daily variations that would significantly impact the data collection. The key temporal requirement was for the soil sampling to coincide with related drum removal activities so that field engineers and the decision maker could evaluate results and address contingencies efficiently.

Define the Scale of Decision Making —The team determined that the relatively small size of the Phase 1 area of concern, together with the preferred soil removal technology, allowed for a practical and cost-effective option to remediate the entire area containing the drums and stained soil. Therefore, they decided to designate the entire Phase 1 area of concern as the scale of decision making. The team agreed that a different approach to the scale of decision making might be appropriate in Phase 2.

Identify Any Practical Constraints on Data Collection — The team determined that no sampling would be conducted on or under existing pavement. The team acknowledged that inclement weather might delay the sampling schedule. In addition, testing with EnviroGard™ should be performed at temperatures between 18 and 27°C (64 and 81°F) for optimal results, and kit materials should be allowed to adjust to ambient temperature. Any sampling data that would be used to make the decision on removing soil would need to occur prior to or during the removal of the drums. Because the team had agreed to use the immunoassay test kits (e.g., EnviroGard™), the number of samples would be constrained by the number of tests per kit for maximum cost-effectiveness. Each kit contains 20 tubes, which the team has decided to run as 16 tests with the other tubes used for quality control samples. Therefore, the team would collect samples in multiples of 16. The potential problem of differing test kit performances due to soil type and moisture content would be examined after collection of the data to determine if significant bias was presented by the chemist and geologist.

Given that the objective was to identify worst-case conditions, and the target population was defined as the surface soil that was visibly stained, the team agreed to use staff expertise to determine the sampling locations, instead of implementing a probability-based sampling design. The team recognized that this application of judgmental sampling—where the subjective selection of sampling locations is based on historical information, visual inspection, or best professional judgment of the sampling team—was valid for this early investigation because the Phase 1 objective was to identify and confirm acutely hazardous conditions.[4] However, the team also recognized that their results and conclusions would have limitations. They knew that their measurements could not be extrapolated beyond the immediate locations at which samples were

[4]USEPA. 1991. *Removal Program Representative Sampling Guidance, Volume 1-Soil*. PB92-963408. Office of Emergency and Remedial Response, Washington, DC.

taken, and that the assumptions underlying most statistical procedures would not hold. These limitations were acceptable to the team because of the limited objective of Phase 1. Regardless of the outcome of Phase 1, the need for defensible conclusions about larger areas, based on valid statistical inferences, would be addressed in Phase 2.

Step 5: Develop a Decision Rule

In this step, all the information is reduced to an "if...then..." statistical decision rule that defines the choice of actions for the decision maker.

Specify the Statistical Parameter that Characterizes the Population of Interest — The contamination scenario describes a problem in which some areas may have much greater levels of Chlordane than the surrounding areas, as evidenced by the visible stains around the drums. Given this "hot spot" contamination scenario, and the team's desire to find the worst areas as part of a screening effort, the team selected the maximum concentration of Chlordane in surface soil as the statistical parameter.

Specify the Action Level for the Decision — The team reviewed readily available state and federal regulatory requirements when establishing an action level for Chlordane. The State of Florida regulatory clean-up level was found to be 4.0 ppm for Chlordane-contaminated soil. The U.S. EPA soil screening level (SSL) for Chlordane was calculated to be slightly higher, at 4.7 ppm SSL for ingestion of Chlordane (a non-carcinogen) in residential soil. The team agreed to use the slightly more stringent state clean-up level of 4.0 ppm as the action level, to ensure that areas known to be above the clean-up level were remediated as part of the removal effort. The team acknowledged that this was a judgment call based as much on potential community perceptions as it was on risk management considerations.

Confirm that the Action Level Exceeds Measurement Detection Limits — The action level of 4.0 ppm exceeds the lower limit of quantitation (LLQ) of SW846 Draft Method 4041, Soil Screening for Chlordane by Immunoassay. Specifically, the test range spans from a lower limit of quantitation (LLQ) of 20 ppb, up to an unspecified maximum screening level.

Combine the Outputs from the Previous DQO Steps and Develop a Decision Rule — The team incorporated the statistical parameter that characterizes the population of interest, the scale of decision making, and the action level. They stated an operational decision rule that would clarify what action to take based on the results of each sample:

> If any one surface soil sample result from an area of visibly stained soil indicates a concentration of Chlordane above 4 ppm, then remove at least the top 6 inches of soil in the Phase 1 area of concern; otherwise do not remove the soil.

Step 6: Specify Limits on Decision Errors

The team already had agreed to use professional judgment to identify visibly stained soils to sample and measure using field tests kits, hence the field sampling would not involve a probability-based sampling design. Consequently, there is no probability-based theory for reliably estimating the magnitude of sampling errors, and any inferences would be confined to the sample locations judgmentally selected in the field. Nonetheless, the team recognized that it was still possible to commit decision errors. Measurement errors could occur during sample analysis. Sampling errors are caused by variability of Chlordane concentrations in the visibly stained soil areas. In the case of a judgmental design, the magnitude of sampling errors can not be reliably estimated, although measurement error can be quantified.

The team identified the two possible decision errors that they could make based on the environmental data: 1) deciding that a visibly stained area is not contaminated with Chlordane at or above 4 ppm when, in fact, the area is contaminated at or above 4 ppm; or 2) deciding that a visibly stained area is contaminated with Chlordane at or above 4 ppm, when in fact, it is not. In the first case, unacceptable contamination would be left on-site, and in the second case, unneeded remediation would be carried out. Next, the team defined the null hypothesis and the alternate hypothesis.

H_o = visibly stained area is contaminated at or above 4 ppm
H_a = visibly stained area is not contaminated above 4 ppm

Once the null hypothesis was stated, the team identified the first decision error described above as a Type I or false rejection error which occurs when the decision maker erroneously rejects the null hypothesis. A Type II error or false acceptance would occur when the decision maker erroneously fails to reject the null hypothesis. The types of decision errors and consequences of those errors are summarized in the table below.

Test Result	**True Value**	**Decision Error**	**Tolerable Decision Error Rate**	**Consequences**
< 4 ppm	≥ 4 ppm	False Rejection: Test result is below 4 ppm and remediation is not needed when, in fact, maximum Chlordane concentration is equal to or above 4 ppm.	*The team did not set tolerable decision error rates because, the judgmental sampling approach does not allow for the assessment of whether or not specific decision error rate limits have been attained.*	Threats to human health and the environment.
≥ 4 ppm	< 4 ppm	False Acceptance: Test result is equal to or above 4 ppm and soil remediation is called for when, in fact, maximum Chlordane concentration is below 4 ppm.		Unnecessary expenditures for further investigation and/or remediation.

The team recognized that it would not be possible to assess whether or not specific decision error rate limits were attained. However, they did obtain quantitative data on measurement method performance from the EPA web site and from Strategic Diagnostics, Inc. (SDI), the company that sells EnviroGard™ Chlordane in Soil Test Kits. The team viewed this information about measurement method performance as a "best case" lower limit on the overall decision error rate (i.e., without being able to know the decision error rate, they knew it had to be greater than the measurement method error rate. In a field trial, 32 soil samples were evaluated by Method 4041 and Method 8080, a well-established laboratory method, at action levels of 1 ppm and 10 ppm. Interpretation of results at 1 ppm resulted in 2 (6.3%) erroneous negative results and 0 (0%) erroneous positive results. Interpretation of results at 10 ppm resulted in 0 (0%) erroneous negative results and 2 (6.3%) erroneous positive results.[5] When the team reviewed this data, they noted that the erroneous results occurred when the true value (according to Method 8080) was near the action level. In addition, the team was aware that immunoassay screening methods often have a positive bias to protect against erroneous negative results (i.e., missing contamination when it is truly there above the threshold). According to SDI, EnviroGard™ Chlordane in Soil Test Kits have a 30% positive bias. The team decided that the overall performance of the measurement method was satisfactory.

Step 7: Optimize the Design

What was unusual in Phase 1 was that the team had discussed the design well before they reached Step 7 due to time constraints and the circumstances of this early Phase 1 investigation. They had agreed to use the immunoassay test kit with a judgmental sampling approach (i.e., nonprobabilistic sampling). The team recognized that choosing a judgmental sampling design instead of a probabilistic sampling scheme would mean that their conclusions would be limited to the immediate vicinity in which a physical sample had been collected, based on the visual staining. The main question to be resolved in Step 7, given all the foregoing requirements and constraints described in Steps 1 through 6, was how to implement the judgmental design in the field so that there was a defensible protocol for selecting the sampling locations and making timely decisions based on the results obtained.

The team reviewed documentation from other sites where Method 4041 had been used as a screening tool. The intended use of Method 4041 at Acconada was consistent with this historical information. In addition, the test kit manufacturers confirmed that the Phase 1 scenario was an appropriate use of the test kit given that the team had existing information on the COC and was prepared to conduct laboratory confirmation of the results. Laboratory confirmation is necessary because of cross-reactivity as described in Step 3. Laboratory confirmation of the results generated by Method 4041 would be conducted as well as laboratory analysis of samples from the drums as Method 4041 cannot test pure product.

---

[5]USEPA. December, 1996. *Method 4041, Soil Screening for Chlordane by Immunoassay*.

The team discussed the issue of sample support because of their interest in obtaining reliable estimates of the true concentration of Chlordane in the collected soil samples. Research has indicated that a larger number of 1-gram or 10-gram aliquots of soil are required to estimate the true concentration of a field sample with specified accuracy as compared to 100-gram aliquots.[6] Although this research addressed multiple aliquots drawn from a larger mass of soil, the basic concept that homogeneity of a soil sample increases as the number of particles increases and/or the volume of individual particles decreases is relevant. Given this information and that the selected test kit requires a 10-gram soil sample, the team agreed to take 100-gram soil samples that would be scooped from the top 2 cm of soil. Each 100-gram soil sample would be homogenized. Then, using a standard subsampling procedure, a 10-gram aliquot would be obtained and prepared for method extraction in accordance with Method 4041. The team reasoned that a 10-gram aliquot of a 100-gram homogenized sample would be more representative of a particular area of stained soil than a 10-gram sample taken directly from that area of stained soil.

Before the test kit could be used, the team had to determine the appropriate dilution factor based on the 4 ppm Chlordane action level. The EnviroGard™ Test Kit includes standard calibrators of 20, 100, and 600 ppb which would not permit a comparison at 4 ppm. Therefore, the team determined that the sample assay would be diluted by a factor of 40, which would allow for comparison at .8, 4, and 24 ppm Chlordane. This is the appropriate dilution because the action level is set in the center of the working range where the precision of an immunoassay is best. In this case, imprecision will increase as the concentration either increases or decreases from 4 ppm Chlordane.

The team discussed that the selected method and action level would require the analysis of an aliquot of a diluted sample and how that dilution may affect the analytical results. For the proposed method, dilution occurs after extraction of the entire 10-gram soil sample. Because an aliquot is drawn from the extract, not the 10-gram sample of soil, the homogeneity of the sample extract is expected to be higher than that of the soil sample. As a result of this discussion, the team agreed that analyzing aliquots of the diluted sample extract, in accordance with the method, should result in acceptable method performance.

In light of the documentation from other sites, EPA, and the manufacturer of the test kits; current knowledge of Acconada site conditions; understanding of the limitations of a judgmental sampling design and Method 4041 performance; and the planned laboratory confirmation, the team agreed that a judgmental sampling design using Method 4041 would be adequate for Phase 1 of the investigation. Therefore, the sampling team proceeded to identify approximately 18 areas of visibly stained soil. They estimated that one sample would be taken from each of the 18 identified areas. However, the team expected that two or more samples would be taken from the areas that appeared to be less homogeneous in appearance, and the assay would be performed in

---

[6]Gilbert, R.O. and P.G. Doctor. 1985. "Determining the Number and Size of Soil Aliquots for Assessing Particulate Contaminant Concentrations." *Journal of Environmental Quality*, 14:286-292.

duplicate to increase the precision of the test. The team anticipated that in addition to the planned samples and necessary QC samples, they may need to retest areas with ambiguous results or sample other areas of interest that are identified during the sampling event. The team agreed to order 4 test kits, which would allow for 64 samples. The results from the test kits would be used to make removal decisions on-site. In addition, all samples would be confirmed by laboratory analysis. While the confirmation data would not be available until after the removal effort had been completed, this data would provide a measure of the test kit performance and would provide additional information on site contamination for Phase 2 of this investigation.

The team ordered the test kits, finalized the DQO outputs, and documented key discussions and assumptions. This information was a critical input for the next activity leading to the Phase 1 data collection, the development of the QA Project Plan.

EPILOGUE

The sampling and analysis described in Step 7 was completed in accordance with the QA Project Plan. The results indicated that most of the areas of visibly stained soil had positive results for Chlordane. Based on these results, the top six inches of contaminated soil was removed with the drums located in front of Building A in accordance with RCRA and CERCLA requirements. A visual inspection indicated no remaining contamination. One additional judgmental sample was taken in each area where the drums had been located, and no Chlordane was detected. Further confirmation sampling was left for Phase 2.

Once the removal action was completed, the planning team began to address the potential for unacceptable contamination in the larger area surrounding Building A. The planning team agreed to work through the DQO Process again as they began Phase 2 of the site assessment. The team anticipated that some of the DQO outputs would remain the same as those for the removal action, but expected that the additional information that they had would be used to refine the outputs for the current iteration. In order to avoid duplication, the outputs that are not changed in a substantive way from the previous outputs are simply summarized below. However, outputs that are significantly different are explained in more detail.

Step 1: State the Problem

Planning Team and CSM — The team and decision maker (i.e., the Acconada manager) remain the same as in Phase 1. The CSM (Figure C-2) is applicable for Phase 2, except that the team was concerned about the possible release of contaminants other than Chlordane, given the possibility that drums containing other contaminants may have been stored temporarily in other parts of the Phase 2 area in the past. The soil and drums in front of Building A were removed based on the analytical results of Phase 1, which indicated that the area had been contaminated with Chlordane. It is possible that all of the unacceptably contaminated soil has not been removed and a threat to human health and the environment remains.

The team anticipated that much of the contaminated soil was removed during Phase 1. However, the team recognized that there would be a potential for contamination in the areas outside the Phase 1 area of concern as well as unacceptable contamination that was not detected in Phase 1. While there are no reports about waste containers being stored to the side or behind Building A, the team considered it a possibility. Furthermore, containers that were stored to the sides or the back of Building A may have leaked or spilled prior to removal. Working from this scenario, the team agreed that they would search for anomalous areas of contamination (i.e., hot spots) that were distinct from the general area around Building A.

Define the Exposure Scenarios — The future use scenario (i.e., mixed-use), potential receptors (i.e., primarily children through dermal exposure or ingestion), and current exposure scenario (i.e., on-site personnel and visitors through direct physical contact with soil or by inhalation of particles) remained the same as in Phase 1.

Specify Available Resources and Constraints — Cost of investigating and remediating the site for a residential future use scenario should be less than the value of the property. Otherwise, other land use alternatives would be considered before significant resources are committed for data collection or remediation. No practical constraints were identified for sampling or analysis. A 6-month target was set for resolving the contamination problem around Building A.

Step 2: Identify the Decision

Principal Study Question — Does site contamination pose an unacceptable risk to human health and the environment? The team defined unacceptable risk in terms of hot spots where stored drums may have spilled or leaked.

Alternative Actions —

- Take a response action, such as remediate the soil, implement institutional or engineering controls to restrict access and potential exposure, and/or recommend further investigation; versus
- Recommend no further evaluation;

Decision Statement —Determine whether site contamination poses an unacceptable risk to human health and the environment and requires further investigation or a response action (e.g., removal, remediation, engineering controls), or recommend that no further investigation is needed.

Step 3: Identify Inputs to the Decision

Information Required to Resolve the Decision — The team reviewed the information that already had been collected (i.e., documentation of site activities, reports from the site workers, and photographs of the site). In addition, the team reviewed the documentation from the Phase 1

data collection (e.g., DQOs, QA Project Plan, analytical results, remedial reports). The team agreed that new environmental data would be required to draw some conclusions about the other areas around Building A.

Information Sources —The environmental data would need to be generated through sampling and analysis of soil around Building A. However, the other information was already available to the planning team. The risk assessor reviewed the Phase 1 data (extent of contamination as well as concentration) and the future use scenario. The risk assessor's primary interest was to identify any areas where waste containers had been stored and if spills or leaks had resulted.

Confirm Appropriate Analytical Methods Exist— Method 8081 (gas chromatography) or Method 8270 (gas chromatography/mass spectrometry) can be used to determine the concentration of Chlordane in soil. The team did not anticipate that Method 4041 would be used as it had been in Phase 1 because the team was less knowledgeable about the nature of contamination in the larger Phase 2 area of concern. Therefore, laboratory methods would be more appropriate should a more complex mixture of contaminants be present.

Step 4: Define the Study Boundaries

Characteristics of the Population of Interest — The team agreed that the Phase 2 investigation would focus on surface soil around Building A, whether or not it was visibly stained. Therefore, the target population was defined as the top 2 cm of soil. The characteristic of interest was the concentration of Chlordane in the surface soil. However, the team had recognized in Step 1 that other contaminants may be present if drums containing contaminants other than Chlordane has been stored in the Phase 2 area of concern.

Spatial Boundaries — The team defined the Phase 2 area of concern as the fenced area surrounding Building A, but excluding the paved access area (Figure C-4). The paved areas are impractical to sample and no significant contamination is expected under the pavement as the area has been paved for several decades. The Phase 2 area is approximately 270 ft x 150 ft (approximately 40,000 ft^2). However, the building and paved area accounted for approximately 12,000 ft^2. Therefore, the Phase 2 area of concern was approximately 28,000 ft^2. The team had no existing data to use as a basis for further subdividing the site, nor any indication that further subdivision was appropriate.

Temporal Boundaries —The team made no specification on when to collect data, nor were there any concerns about cyclical phenomena that might affect the sampling and analysis.

Size and Intensity of Hot Spots — The data from Phase 1 and the site history indicated that drums were typically stored in clusters of four, the number of drums that would fit on a single pallet. If four drums were placed on a pallet, they could be contained in a circle with an approximate diameter of 10 feet. The Phase 1 data also indicated that a single positive

measurement of Chlordane at or above 4 ppm (the action level in Phase 1) was sufficient to identify an area of concern. Therefore, the risk assessor decided that the hot spots in Phase 2 could be reliably observed by a single measurement at or above 4 ppm Chlordane.

Practical Constraints on Data Collection — No sampling would be done under existing pavement. Inclement weather could affect the sampling schedule.

Step 5: Develop a Decision Rule

Confirm that Measurement Detection Limits are Appropriate — Method 8081 (gas chromatography) or 8270 (gas chromatography/mass spectrometry) are appropriate for sampling Chlordane and related chemicals. The detection limit for both methods is well below 4 ppm.

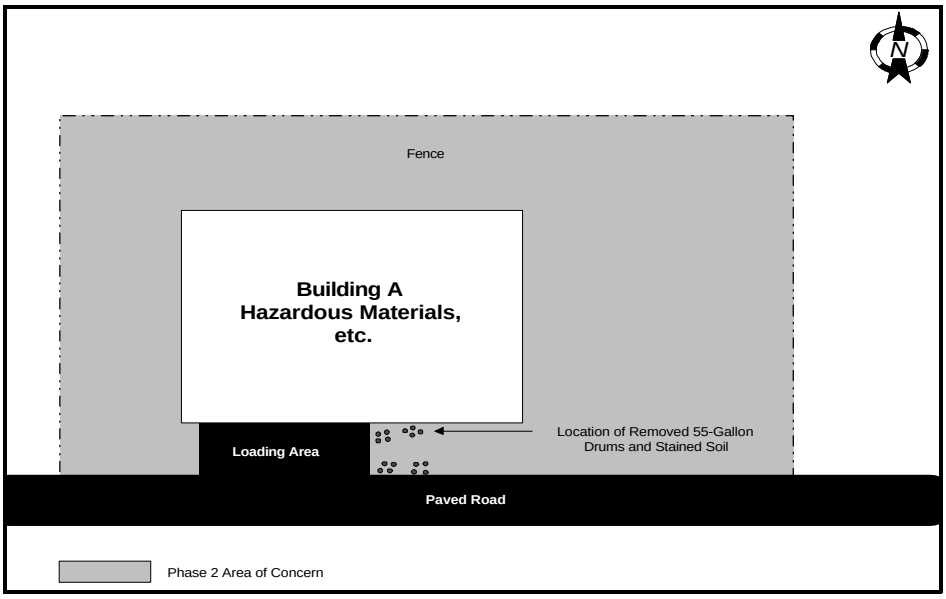


Figure C-4. Building A and Surrounding Area - Phase 2

Decision Rule — If at least one hot spot with a diameter of 10 ft or greater and at least a 4 ppm Chlordane concentration exists, then investigate the boundaries of the hot spot; otherwise conclude that the Phase 2 area does not require remediation.

Step 6: Specify Limits on Decision Errors

Determine the Possible Range of the Parameter of Interest — The data from Phase 1 indicated that a range of 0-100 ppm Chlordane was appropriate.

Define Both Types of Decision Errors, Consequences, and the Baseline Condition — The

team identified the two possible decision errors that they could make based on the environmental data. Next, they established the consequences of those errors.

Decision Errors, Consequences, and the Baseline Condition —

Decision Error	Consequences
"Decide at least one hot spot with a diameter of 10 ft or greater and at least a 4 ppm Chlordane concentration does not exist and remediation is not necessary when, in fact, a hot spot does exist."	Threats to human health and the environment.
"Decide at least one hot spot with a diameter of 10 ft or greater and at least a 4 ppm Chlordane concentration does exist and remediation is necessary when, in fact, a hot spot does not exist."	Unnecessary expenditures for further investigation and/or remediation.

The first decision error listed in the table above would occur when no single measurement indicated a Chlordane concentration at or above 4 ppm. This decision error could occur as a result of measurement error, or if the hot spot was very heterogeneous, such that some areas within the 10 foot diameter were below 4 ppm, while other areas were at or above 4 ppm. The second decision error listed above (i.e., deciding that at least one hot spot with a diameter of at least 10 ft does exist and further investigation is necessary when, in fact, a hot spot does not exist) could occur in two ways: (a) if an area of elevated concentration *smaller* than the defined hot spot diameter of 10 ft happened to fall on a sampling grid location; or (b) if the laboratory erroneously reports a Chlordane concentration of at least 4 ppm for a measurement at any sampling location.

The team agreed the baseline condition should be "at least one hot spot exists." The hot spot is considered to exist with one positive measurement for Chlordane at or above 4 ppm. Therefore, the null and alternate hypothesis can be stated as:

H_o = at least one hot spot with a diameter of 10 ft or greater and at least a 4 ppm Chlordane concentration exists.
H_a = at least one hot spot with a diameter of 10 ft or greater and at least a 4 ppm Chlordane concentration does not exist.

The baseline condition establishes which of the decision errors described above is a false rejection (Type I) error and which is the false acceptance (Type II) error. A false rejection error occurs when the decision maker rejects the null hypothesis in favor of the alternate hypothesis based on the observation of misleading environmental data. In other words, a false rejection error would occur if the decision maker decides that a hot spot does not exist when, in fact, it does. Conversely, a false acceptance decision error occurs when the decision maker incorrectly fails to reject the null hypothesis (i.e., when the decision maker decides that a hot spot does exist when, in

fact, it does not.)

Gray Region — No grey region is established for this phase of the study because there are only two possible outcomes: a hot spot with a diameter of 10 feet or greater and at least a 4 ppm Chlordane concentration exists or a hot spot does not exist.

Tolerable Probability for Decision Errors — The team agreed that a 0.80 probability of hitting a round hot-spot with a diameter of 10 feet or greater and at least a 4 ppm Chlordane concentration was acceptable.

Step 7: Optimize the Design

The team wanted to be able to draw conclusions about the entire area that was sampled and not just the precise sample locations. As a result, the team did not consider a judgmental sampling scheme for this iteration as they had in Phase 1. The team evaluated a number of probabilistic sampling schemes. The team determined a grid design to be the most appropriate design that would meet the specified DQOs.

Grid sampling uses a specified pattern (e.g., square, triangular, rectangular, or hexagonal grid) along which samples are taken at regular intervals. The location of the first sample is chosen at random and the remaining (n-1) sampling locations are placed according to the specified pattern. The advantage of grid sampling is that the target population is uniformly represented in the sample. In addition, grid sampling is practical to implement in the field. Grid sampling is commonly used when searching for hot spots. One disadvantage of grid sampling is the possibility that the grid will be aligned with some existing pattern of contamination.

A square grid was selected over other grid shapes because it is simpler to implement in the field. Grid spacing is determined by the shape and size of hot spots as well as the desired confidence of locating a hot spot of the specified shape and size. The smaller a hot spot one is trying to find, the more dense the grid needs to be. An elliptically-shaped hot spot requires a finer grid than a circular shape. As discussed previously, the hot spots were expected to be circular and at least 10 ft in diameter. The team selected a sampling design that would detect a round hot-spot (diameter 10 ft or greater) with 0.80 probability. The team determined that a 9.9 foot square grid would be required to meet the team's DQOs. This would require 285 samples locations to cover the approximately 28,000 ft^2 area of concern. The team recognized that a 10 foot grid would be more practical to implement in the field. By increasing the grid size to 10 feet, and maintaining a probability of detection at 0.80, a hot spot diameter of 10.1 feet could be detected using the same number of samples. The planning team decided this was acceptable. The sampling locations for the 10 foot square grid are shown in Figure C-5.

While the team remained concerned about larger hot spots (diameter of 10 feet or greater), they acknowledged that smaller sizes could occur. Therefore, once the team had

selected the grid size which met their constraints, they were interested to know the probability of detecting hot spots of various sizes given the selected 10 foot grid size. The probability of hitting a hot spot (y) of a given diameter (x) is plotted in Figure C-6. Furthermore, the team recognized that these probabilities were somewhat higher than they could expect to achieve in practice because of the somewhat idealized assumptions underlying the standard performance curve for hot spot detection, as in Figure C-6 (e.g., homogeneity of contamination, uniform circular shape of hot spot, measurement system that always detects presence of hot spots). After discussing these issues, the team remained in agreement that the 10 foot square grid was an adequate design.

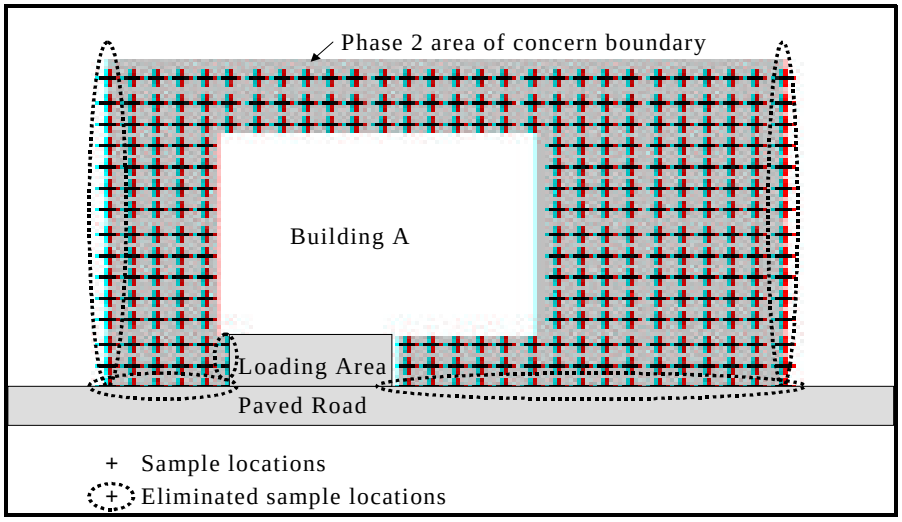


Figure C-5. Phase 2 final sample locations with 52 sample locations eliminated from initial plan

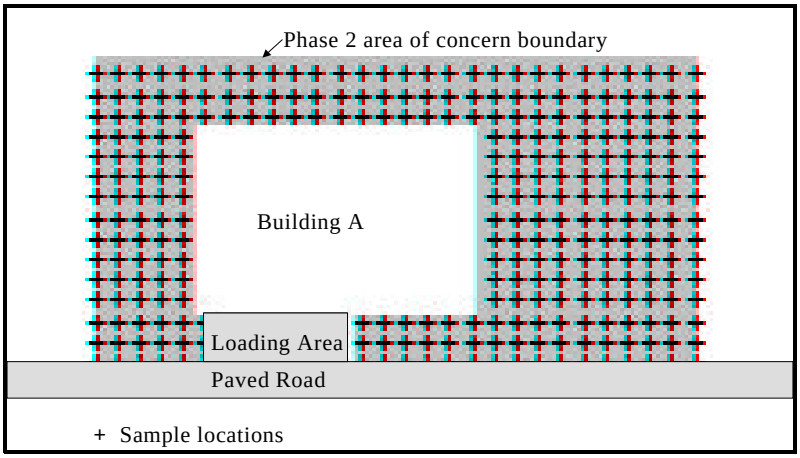


Figure C-6. Phase 2 initial sample locations based on 10 foot square grid

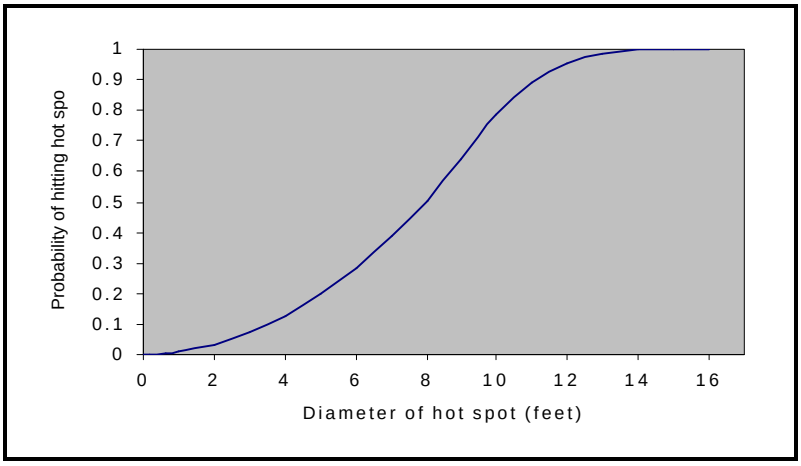


Figure C-7. Probability of hitting a hot spot with a diameter of x feet given a 10 foot square grid

The team considered the sample locations shown in Figure C-5. They noted that along the eastern and western boundaries, sample locations were almost directly on the boundary line. The team choose to eliminate these locations from the design because they had been conservative in establishing the boundary. In addition, along the southern boundary of the area of concern, sample locations were identified just at the edge of the paved loading area and road. The team opted to eliminate these locations from the design as well because the edges of the pavement were not precise and would overlap many of these sample locations. The team did not consider eliminating any sample locations from the northern boundary because the sample locations were a couple feet inside the boundary. In sum, 52 samples were eliminated from the initial square design for a revised total of 233 sample locations for Phase 2. The estimated cost of implementing a square grid design with 233 samples was within the budget.

Conclusion

Prior to Phase 2 sampling the team prepared a QA Project Plan as required, which included documentation of the design illustrated in Figure C-7 and the related assumptions. Phase 2 sampling was conducted in accordance with the QA Project Plan. Analysis of the Phase 2 data indicated Chlordane contamination above 4 ppm near the western side of the building. Based on these results, the team agreed that the next step for this site was to explore the boundaries of the hot spot(s) before any remedial action was taken in a third phase of investigation. Although a description of the planning and implementation of Phase 3 is beyond the scope of this case study, the Phase 3 investigation resulted in the delineation and removal of contaminated soil along the west side of Building A.

APPENDIX D

PROBABILISTIC SAMPLING
DQO CASE STUDY: BLUE MOUNTAIN SMELTER

Background

The Blue Mountain Smelter site (Blue Mountain) is a 150-acre site located in the southeastern United States, approximately one-half mile from the coast of the Gulf of Mexico. The elevation of the site is near sea-level and portions of the site are marshy. Large petrochemical and industrial complexes are located north and northeast of the site. Industrial waste disposal facilities and undeveloped marshy areas neighbor the site to the south and southwest. Residential areas are west and northwest of the site.

The water table is only a few feet below the ground surface. Surface waters surrounding the site are brackish with gradient flow toward the ocean (southeast). Ground water flow follows the gradient, towards the Gulf of Mexico.

Temperatures in this area range from 10-50 °F in winter to 70-110 °F in summer. Prevailing winds are from the northwest and are generally steady from 5-15 mph with gusts to 50-70 mph during frequent summer thunderstorms. Precipitation in the region is 100-150 inches per year with 50 percent falling during the spring months (March through June). The remaining 50-75 inches of annual precipitation are distributed unevenly throughout the remaining months. A site plan depicting points of interest is shown in Figure D-1.

Tin smelting operations began at Blue Mountain in 1941. The site ownership and plant operation changed several times without major restructuring of smelting processes until copper smelting operations were added in 1989.

Operations resulted in the production of a variety of wastes, many of which still remain, untreated and onsite. As of 1992, piles of residual smelting wastes covered approximately 10 acres of the site. Iron-rich liquids were recovered using ponds averaging 10 feet deep and covering about 80 acres of the site. Oxidized ferric chloride collected in the ponds was sold to wastewater treatment operators until 1983, when ferric chloride production ceased at Blue Mountain because of changes in the smelting process. A scrubber system used for removing sulfur dioxide from stack emissions produced calcium sulfate (gypsum) sludge, which was also ponded onsite (Figure D-1).

During the 1970s and 1980s, spent catalysts were stored onsite with minimal recovery efforts. Some uranium-bearing spent catalysts, considered to be low-level radioactive materials, were buried in a permitted landfill located in the southern part of the site in 1978. This landfill is clearly delineated and monitored quarterly by a state agency.

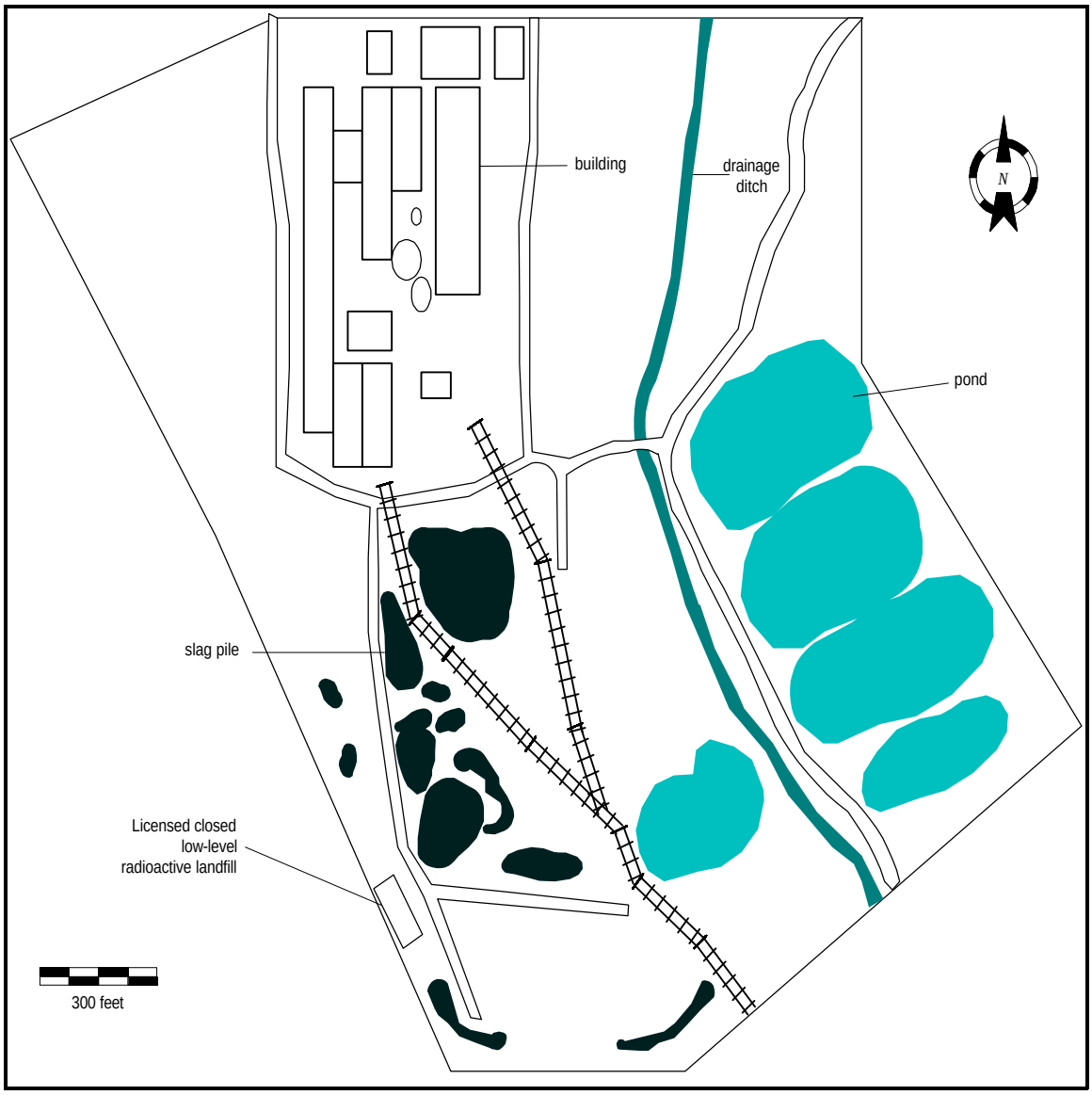


Figure D-1. Blue Mountain Smelter Site Map

In the early 1980s, a small area near the smelter building was leased for the processing of still bottoms and waste oil from chemical and refining companies. Several associated tanks and drums and a small building remained at the time this study was conducted. It is estimated that buildings cover 15 acres of the site. At the time the DQO Process was initiated, wastes remaining from site activities had not yet been adequately characterized.

The Blue Mountain Smelter site was listed in the Comprehensive Emergency Response, Compensation, and Liability Information System in 1979. Based on its Hazard Ranking System (HRS) score, this site was proposed to be added to the National Priorities List (NPL) in 1988, and was placed on the NPL in 1990. A neighboring industry purchased the onsite slag piles for

recovery of metals, and thereby relieved the potentially responsible party (PRP) of removal and disposal obligations for the slag wastes.

Site characterization efforts included early assessment remedial investigations for which limited amounts of data were available for planning. Preliminary onsite surface soil analyses indicated concentrations of arsenic as high as 720 mg/kg; some analyses also indicated above background levels of cadmium, copper, and mercury.

Because limited resources were available for this effort, the U.S. Environmental Protection Agency (EPA) negotiated the consent of all stakeholders to treat this site as a pilot for the Superfund Accelerated Cleanup Model (SACM). Based on earlier site work, the treatment decision was classified as an Advanced Assessment Decision, Phase I.

Step 1: State the Problem

This step summarizes the contamination problem, identifies members of the planning team, develops a conceptual site model, and identifies exposure scenarios.

Identify Members of the Regional Decision Team-The planning or Regional Decision Team (RDT) was led by the EPA Regional Remedial Project Manager (RPM) together with a chemist from the EPA Regional Environmental Services Division (ESD), a risk assessor from the EPA Regional Superfund office, a representative of the firm contracted by the PRP to conduct remediation activities, a hydrogeologist, the EPA Regional Superfund Quality Assurance Officer, a soil scientist with statistical training, and the site project manager representing the PRP.

Develop the Conceptual Site Model-Figure D-2 depicts the Conceptual Site Model (CSM), which links the primary and secondary sources of contamination, the mechanisms of release to the environment, the exposure pathways, and exposure routes to the receptors. After reviewing the CSM, the RDT identified soil contamination as the most critical issue that was not already being addressed. Therefore, for this DQO Process, they limited their focus to surface soil. Other media were being addressed under separate efforts and were, therefore, not a concern to the RDT for this investigation.

The RDT listed residual metals from the smelting operations as the primary contaminants of concern (COCs) at this site. Sampling efforts undertaken during previous site investigation activities revealed levels of arsenic as high as 720 mg/kg. In addition, elevated levels of cadmium, copper, and mercury had been found in some samples. At the time of this study, no applicable or relevant and appropriate requirements (ARARs) existed for surface soils contaminated with heavy metals in an industrial setting.

Define Preliminary Exposure Scenarios-The RDT believed that sources of actual and potential contamination were contaminants lying within six inches of the surface of the soil, in ponds, in the landfill, in tanks and drums, and in visually identifiable slag piles. The RDT assumed

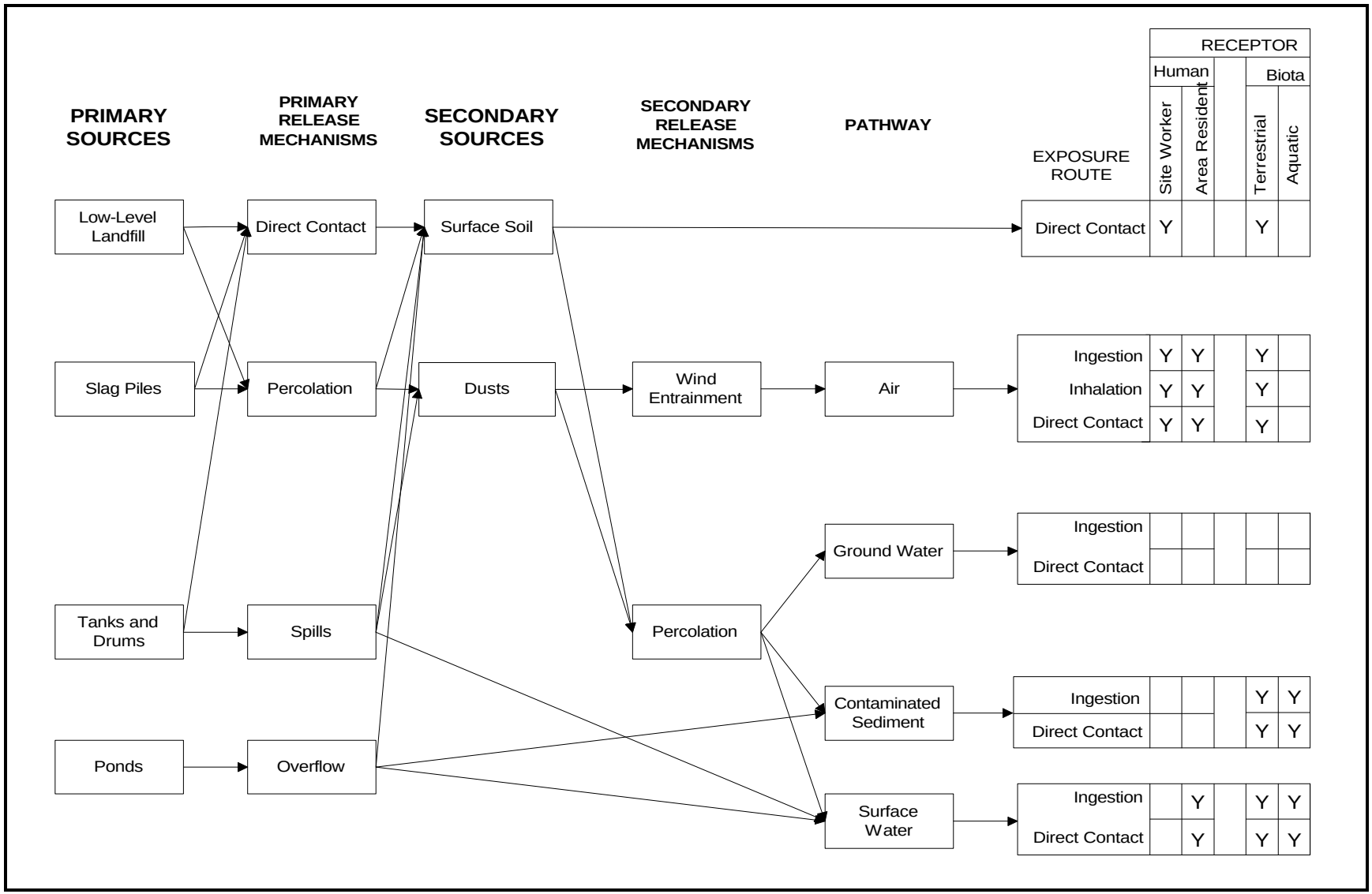


Figure D-2 Conceptual Site Model

an industrial future land use scenario for this site. Based on the history of land use of the site and surrounding area, residential and recreational future use scenarios seemed unreasonable.

The primary exposure scenario at this site was determined to be heavy metal contamination of the surface soils ingested or inhaled by onsite workers through wind-entrained dusts (stirred up by direct contact). All other exposure scenarios (e.g. airborne heavy metal exposures to offsite residents and biota) were secondary to the primary exposure scenario and were being addressed through other efforts.

Specify Available Resources and Constraints-The RDT's funding for this study, which was provided by the PRP, was approximately $50,000. Based on the results of this study, the team would address and finance subsequent work, such as the remedial design, with additional funding.

Site workers and local residents living west and northwest of the site were concerned about the chronic effects of the airborne dust and direct exposure to the contaminants of concern. A practical limit of 6-12 months for completion of this study was well received in discussions with residential and employee groups.

Step 2: Identify the Decision

This step requires the team to identify the decision that will address the contamination problem.

Identify the Principal Study Question-As other primary sources of contamination were either being removed or were being addressed in separate investigations, the RDT confined the scope of their study to onsite surface soil. The RDT believed that contaminated soil posed a threat to the environment, primarily through ingestion and inhalation of wind-entrained dusts caused by direct contact and, secondarily, through leaching of contaminants to ground water and contaminant runoff to surface waters and sediments. The RDT identified as the following principal study question: "Do the concentrations of heavy metals in the surface soil exceed risk-based concentration limits?"

Identify Alternative Actions-The alternative actions that could result from the resolution of the principal study question are: Recommend site status to be listed as Site Evaluation Accomplished (SEA), or recommend further assessment or a possible response action.

Combine the Principal Study Question and the Alternative Actions into a Decision Statement-The first action eliminates the need for further study or cleanup activities at this site, the second necessitates additional assessment and/or cleanup work. The RDT combined these two outcomes with the study question to formulate the decision statement:

Determine whether heavy metal contamination of the surface soil poses a hazard to worker health by exceeding risk-based concentration levels and warrants remediation, or whether the contamination is less than the risk-based concentration levels and investigators may proceed with a Site Evaluation Accomplished (SEA) determination.

Step 3: Identify Inputs to the Decision

In this step, the RDT identifies the types of information needed to resolve the decision statement and the method for obtaining this information. The RDT identifies the information required to resolve the decision statement, as well as the sources for each informational input. The RDT then determines which health-based or risk-based criteria should be used to determine the action level. Finally, the appropriate measurement methods that will provide the necessary data are identified.

Identify the Information Required to Resolve the Decision Statement-In order to determine whether concentrations of metals in the surface soils exceeded risk-based concentration limits the RDT needed to answer three questions:

- What types of heavy metals are present in the surface soils?
- What EPA human health risk measures exist to assess potential worker health risks?
- Does the surface soil pose a hazard to worker health?

The RDT relied upon the Conceptual Site Model and site-specific risk assessment to develop Preliminary Remediation Goals (PRG) for the contaminants that present the most risk to onsite workers. Although previously collected data would be used for initial estimates of contaminant distribution and maximum contaminant concentrations, the RDT determined that they would need to collect new environmental measurements to adequately resolve the whether the surface soil poses a risk to worker health.

Determine The Sources For Each Informational Input-The RDT examined all of the previously conducted surface soil studies at the site and found that four contaminants had been observed at concentrations above background levels: arsenic, cadmium, copper, and mercury. Contaminant toxicity values for these four metals were gathered from the Integrated Risk Information System (IRIS) and Health Effects Assessment Summary Tables (HEAST) for both carcinogenic and noncarcinogenic effects (Table D-1). The RDT then performed a concentration toxicity screen for these contaminants, as suggested by Superfund DQO guidance (EPA, 1995). The results of this toxicity screen (Table D-2) indicated that 99 percent of total risk to workers is due to arsenic in the surface soils. Thus, the RDT narrowed the list of COCs to only one contaminant, arsenic.

Table D-1. Toxicity Information for Contaminants of Concern at Blue Mountain Smelter Site

NONCARCINOGENIC TOXICITY

	Oral		Inhalation	
Constituent	**RfD_0 (mg/kg-day)**	**Source**	**RfD_i (mg/kg-day)**	**Source**
Arsenic	3.0E-04	IRIS	--	--
Cadmium	1.0E-03	IRIS	--	--
Copper	3.7E-02	HEAST	--	--
Mercury	3.0E-04	HEAST	3.0E-04	IRIS

CARCINOGENIC TOXICITY

	Oral			Inhalation		
Constituent	**SF_0 $(mg/kg/day)^{-1}$**	**Wt of Evidence**	**Source**	**Sf_i $(mg/kg\text{-}day)^{-1}$**	**Wt of Evidence**	**Source**
Arsenic	1.5E+00	A	IRIS	1.5E+01	A	IRIS
Cadmium	--	B1	EPA_ED10	6.3E+00	B1	IRIS
Copper	--	--	--	--	--	--
Mercury	--	--	--	--	--	--

Note: Wt of evidence rankings are based upon EPA Cancer Guidelines, which define Group A and Group B toxins as the following:

Group A: Known human carcinogen. Sufficient epidemiological evidence to support casual association between exposure and cancer; and

Group B: Probable human carcinogen. Limited evidence in epidemiologic studies (B1) and/or sufficient evidence from animal studies (B2).

SF = cancer slope factor

RfD = reference dose

-- : these data had not been developed by EPA at the time of publication.

Table D-2. Concentration Toxicity Screen for Contaminants of Concern at the Blue Mountain Smelter Site

	Noncarcinogenic Contaminant Toxicity		Carcinogenic Contaminant Toxicity		Onsite Soils		
Constituent	**Oral RfD_0 (mg/kg-day)**	**Inhalation RfD_i (mg/kg-day)**	**Oral SF_0 $(mg/kg\text{-}day)^{-1}$**	**Inhalation Sf_i $(mg/kg\text{-}day)^{-1}$**	**Max. Conc. (mg/kg)**	**Risk Factor**	**Percent of Total Risk**
Arsenic	3.0E-04	—	1.5E+00	1.5E+01	720.00	2.E+06	99
Cadmium	5.0E-04	—	—	6.3E+00	.94	2.E+03	<1
Copper	3.7E-02	—	--	—	130.00	4.E+03	<1
Mercury	3.0E-04	8.6E-05	—	--	.18	6.E+02	~0

Note: RfD = reference dose

SF = cancer slope factor

-- = data not available

Identify The Information Needed to Establish The Action Level-At the time this study was conducted, EPA's policy was to remediate to a risk-based cleanup level between 10^{-6} to 10^{-4} excess risk of cancer and a noncarcinogenic hazard quotient (HQ) less than 1.0, as specified in the preamble to the National Contingency Plan (40 CFR 300.430(e)(2)). A PRG for arsenic was developed by the RDT to meet these risk-based levels. In accordance with *Risk Assessment Guidance for Superfund: Volume I—Human Health Evaluation Manual (Part B, Development of Risk-based Preliminary Remediation Goals, EPA/540/R-2/003*) (EPA, 1991), the following site-specific data were gathered: media of concern (surface soil), chemical of concern (arsenic), and probable future land use scenario (industrial). The RDT employed a site-specific risk assessment to develop a PRG that met these criteria.

Confirm That Appropriate Measurement Methods Exist-The RDT selected SW-846 method 7060 as the most appropriate analytical method for this investigation, providing the most accuracy and precision available for measuring arsenic in soils. Table D-3 presents a summary of characteristics of the 7060 method.

Table D-3. The Selected SW-846 Analytical Method for Measuring Arsenic in Surface Soil

Analyte	Method	Equiv. Method	Principle	Bias (R%)	Precision (RSD%)	MDL (mg/kg)	Cost ($/analysis)
Arsenic	7060	206.2	GFAA[1]	96	5	1	75

[1]GFAA = Graphite Furnace Atomic Absorption

Step 4: Define the Boundaries of the Study

In this step, the RDT further defines the limitations and interpretations of the DQO analysis, determines the geographic and temporal boundaries and identifies economic and practical constraints.

Identify the Spatial Boundaries-The site boundaries were selected as the geographical boundaries for this study. Because arsenic is generally water-insoluble and does not tend to migrate in soil, the RDT believed that contamination would not be likely to spread offsite, even within a lengthy sampling and analysis time frame. The general stability of arsenic in soil gave the RDT flexibility in planning sampling and analysis.

The spatial boundaries were based on concern over long-term exposure to workers. Although the depth of contamination was not known, the RDT decided that for initial planning purposes the top 6 inches of soil was the limit to which onsite workers could be exposed. Although the RDT examined the possibility of worker exposure to soils deeper than 6 inches during construction or excavation activities, the RDT decided that these exposures would generally be short-term and would not pose a threat to worker health.

Specify the Scale of Decision-making-The scale of decision-making is defined as the smallest unit to which the decision rule is applied. The goal of the RDT was to establish a scale of

decision-making that minimized total costs of site investigation (planning, sampling, and analysis) as well as remediation. The RDT wanted to balance the cost of sampling many small units, which would require taking many samples, with the cost of taking fewer samples by delineating larger sampling units. With smaller units the RDT would reduce total remediation costs of the site's surface soil by cleaning up smaller contaminated areas. With larger areas the RDT would reduce sampling and analysis costs by having fewer samples collected and analyzed.

The RDT decided to divide the site into EUs. An EU is the expected area over which an individual may be exposed to contaminated media while performing routine activities during a specified time. The risk assessor assumed that the average time worked at the site was 8 hours a day, 5 days a week over 30 years. Over this time period, the worker spent time in a small area, visiting other areas only occasionally and perhaps never visiting other more remote or inaccessible areas.

The persons most likely to receive the highest doses of contamination were the onsite workers. Based on previous studies of similar sites, one half acre (21,840 sq. ft.) was considered to be the smallest reasonable area covered by an onsite worker during daily industrial activities. Hence, for this soil cleanup effort, EU was defined as 21,840 square feet.

For this 150-acre site, the RDT determined that they needed to address approximately 45 acres of soil in this decision, after excluding buildings (15 acres), ponds (80 acres), and slag piles (10 acres). This area comprised approximately 90 EUs.

Identify the Temporal Boundaries-There was no additional temporal boundary placed on the study since arsenic is relatively stable over time. Therefore, it was not imperative that the RDT investigate and remediate within a short time frame. However, there was considerable worker and public concern over on-going exposure at the site. In order to minimize public concern, the team planned on finishing this study within 6 to 12 months. There were no seasonally induced boundaries on sample collection activities since the climate of the area allowed for year round sampling and arsenic concentrations in the surface soil were not known to fluctuate with climate.

Identify Practical Constraints-Site samples had to be collected during the third plant shift (11:00 p.m. - 7:00 a.m.) to avoid interference with routine daily plant activities. A member of the site owner’s security staff had to be employed to satisfy the owner’s legal and safety concerns.

Step 5: Develop a Decision Rule

In this step, the team combines the qualitative information about site contamination with measurable, health-based concentration criteria in an “if...then..” statement called the decision rule.

Specify the Statistical Parameter Characterizing the Population of Interest-The RDT was more concerned with the chronic health effects of arsenic contamination in the surface soil than with acute effects. In measuring the long-term effects of a heavy metal in the surface soil, risk assessors use the mean concentrations of the COCs, because this parameter best represents the random integration of exposure over the long term. Hence, the RDT selected mean concentration of arsenic within each EU (21,840 sq. ft.) as the most appropriate parameter characterizing the population of interest.

Specify the Action Level for the Decision-The Preliminary Remedial Goal (PRG) that the RDT calculated for arsenic in the surface soil was 600 mg/kg. This PRG was calculated in accordance with *Risk Assessment Guidance for Superfund: Volume I—Human Health Evaluation Manual (Part B, Development of Risk-based Preliminary Remediation Goals, EPA/540/R-2/003)* (EPA, 1991) and compared favorably with soil cleanup levels at other Superfund sites. Soil cleanup concentrations for arsenic have ranged from 70-200 mg/kg for sites with anticipated future residential use to 500-1000 mg/kg for sites with anticipated industrial use.

The RDT selected an arsenic concentration of 600 mg/kg as the action level for this site. The detection limit of the analytical method proposed for this study (see Table D-3) was 600 times lower than the action level selected.

Develop The Decision Rule-The decision rule is as follows:

If the mean concentration of arsenic in the surface soil within an EU is less than 600 mg/kg, then do not study the EU further and consider Site Evaluation Accomplished.

Otherwise, if the mean concentration of arsenic in the surface soil within an EU is greater than or equal to 600 mg/kg, then continue with investigation and/or remediation of this EU.

Step 6: Specify Tolerable Limits on Decision Errors

In this step, the RDT establishes quantitative performance criteria for the sampling design. Tolerable probability values are assigned for each type of potential decision error.

Determine The Possible Range of The Parameter of Interest-The highest soil concentration of arsenic observed at this site in previous investigations was 720 mg/kg (see Table D-2). The RDT selected an arsenic concentration of 2000 mg/kg as a conservative maximum mean concentration within an EU, after considering the known smelting activities at the site. Surface soil arsenic concentrations at previously studied smelter sites have generally been much lower than 2000 mg/kg. The lower limit for arsenic was set at 0 mg/kg, because arsenic is known to occur naturally in low levels in soil.

Define Both Types of Decision Errors and Their Potential Consequences-Two potential decision errors could be made based on interpreting sampling and analytical data:

Decision Error A: Concluding that the mean arsenic concentration within an EU was less than 600 mg/kg when it was truly greater than 600 mg/kg, or

Decision Error B: Concluding that the mean arsenic concentration within an EU was greater than 600 mg/kg when it was truly less than 600 mg/kg.

The consequences of Decision Error A, *incorrectly deciding an EU was "clean" (mean arsenic concentration less than 600 mg/kg),* would have immediate and future health implications, because that EU would be listed as Site Evaluation Accomplished and would not be evaluated further. This decision would leave contaminated soil undetected and would likely increase health risks for onsite workers. Furthermore, future investigations of the site could reveal the true, hazardous level of contamination, which could possibly present legal and credibility problems for the EPA.

The consequences of Decision Error B, *incorrectly deciding an EU was "not clean" (mean arsenic concentration greater than or equal to 600 mg/kg),* would cause the needless expenditure of resources (e.g., funding, time, sampling crew labor, and analytical capacity). As a result, the RDT would be less capable of adequately responding to truly pressing problems at this site (e.g., remediation of ponds and other contaminated areas). Either these needs would not be addressed, or limited EPA resources would be expended in order to complete the additional work. Furthermore, it is likely that the next phase of investigation would reveal the true benign level of contamination, and the EPA could be accused of being overly cautious and wasteful.

After examining the consequences of both decision errors, the RPM decided that Decision Error A, *incorrectly deciding that the mean arsenic concentration is less than the action level of 600 mg/kg,* posed more severe consequences because the true state of soil contamination ([As] > 600 mg/kg) could go undetected for months or even years, all the while exposing onsite workers to unacceptable concentrations of arsenic.

Consequently, the baseline condition chosen for this site was *that the mean arsenic concentration within an EU was truly greater than or equal to the action level of 600 mg/kg.*

In statistical language, the baseline condition becomes the null hypothesis (H_0) and the alternative, the alternative hypothesis (H_a). This can be written as:

H_0: $[\text{arsenic}]_{mean} \geq 600$ mg/kg
H_a: $[\text{arsenic}]_{mean} < 600$ mg/kg

A false rejection decision error occurs when the null hypothesis is falsely rejected. In this case, such an error would have occurred if the RDT decided that the mean was less than 600 mg/kg, when in fact, the true mean soil concentration was greater than or equal to 600 mg/kg.

A false acceptance decision error occurs when the null hypothesis is falsely accepted . In this case, such an error would have occurred if the environmental data indicated that the mean concentration was greater than or equal to 600 mg/kg when, in fact, the true concentration was less than 600 mg/kg.

Specify the Boundaries of the Gray Region-The gray region defines a range that is less than the action limit, but too close to the action limit to be considered "clean," given uncertainty in the data. When the null hypothesis (baseline condition) assumes that the site is contaminated (as in this case study), the upper limit of the gray region is bounded by the action level; the lower limit is determined by the decision maker.

The RDT evaluated the potential of making false acceptance errors (determining incorrectly that further investigation is not needed for the EU) and decided that it was very important not to make false acceptance errors. However, to decrease the likelihood of committing false acceptance errors, the RDT would need greater confidence in the data that were collected, which would require increased sampling and analysis (and increase total cost of the field investigation). Weighing the costs of increased sampling and analysis versus the costs of false acceptance errors, the RDT chose 500 mg/kg as the desired lower limit for the gray region.

The RDT chose 500 mg/kg with the full understanding that this could be subsequently altered depending on what occurred in Step 7 of the DQO Process. The RDT was aware that several iterations of the DQO Process could be necessary before settling on a final design to collect the data.

Assign Probability Values to Decision Errors-Following *The Data Quality Objectives Process, (EPA QA/G-4)* (EPA, 1994), the RDT initially set the allowable decision errors outside the 500-600 mg/kg gray region at 1 percent (p = .01). This means that the RDT wanted to collect and analyze enough samples so that the chance of making either a false rejection or a false acceptance decision error was only one-in-a-hundred, an exceptionally stringent criterion that would demand many samples. The RDT planned to use DEFT [*The Data Quality Objectives Decision Error Feasibility Trials Software, (EPA QA/G-4D*) (EPA, 1994)] to aid with preliminary design in Step 7 and to explore other design options in order to optimize data quality within the given budget of $50,000. The RDT determined that the DEFT capabilities were applicable to this problem and were adequate to support the initial design activities in Step 7. The information collected for this step of the DQO Process is summarized in Table D-4.

Table D-4. Initial Data Quality Criteria

Needed Parameter	Criteria
Action Level	600 mg/kg
Gray Region	500-600 mg/kg
Null Hypothesis (H_0)	Mean [As] $\geq$ 600 mg/kg
False Acceptance Decision Error Limit	chance of decision error = .01 at 500 mg/kg
False Rejection Decision Error Limit	chance of decision error = .01 at 600 mg/kg

Step 7: Optimize the Design for Data Collection

In this final step of the DQO process, the RDT used the DQO criteria that were identified in Steps 1 through 6 to explore the feasibility of various data collection alternatives. This step also allowed the RDT to identify and reject options that did not meet the DQOs (i.e., would not produce data sufficient for the decision quality that was specified). In so doing, it helped the RDT discover which designs did not provide information of acceptable quality.

Review The DQO Outputs And Existing Environmental Data-Two factors drove the RDT's design decision: the cost and the quality of the environmental samples collected and analyzed. Their goal was to gather data of acceptable quality within the specified budget of $50,000. Using the SW-846 analytical method listed in Table D-3, the initial cost estimate for collecting a field sample was $100, with a laboratory analysis cost of $75 per physical sample. The relative standard deviation (rsd) for the measurement was 5% (Table D-3) and so the estimated measurement standard deviation for the method operating at the Action Level of 600 was then 30 (5% X 600). Turning to the field variability component, the RDT decided to consider the "worst case" scenario where the field variability is estimated at 10 times the laboratory variability is the estimated field variability = 300. Combining these variabilities together to create the total variability (total variability = field variability + laboratory variability, where the variability is in the form of the statistical variance) gives total variability $= 300^2 + 30^2$ =90900, giving the estimated total standard deviation as $\sqrt{90900} = 302$.

As the maximum observed value was 720 (Table D-2), this figure was deemed appropriate by the RDT.

Identify General Data Collection Design Alternatives-The RDT considered two main design alternatives: simple random sampling and composite sampling. The RDT planned to first explore was simple random sampling for 90 EUs. The RDT would explore compositing and other options only if the simple random alternative proved too costly.

Formulate The Mathematical Expressions Necessary For Each Data Collection Design

Sample Size

The RDT used DEFT to explore the impacts of the design constraints described in the previous sections. For initial calculations, DEFT offered a simple random sampling strategy using a t-test to calculate sample size. This approach assumed that each sample collected would be analyzed once, and that sampling and analysis variability was uniform for the set of samples considered. The sample size formulas used in the calculations was:

$$n = \frac{\hat{\sigma}^2 (z_{1-\alpha} + z_{1-\beta})^2}{\Delta^2} + \frac{z_{1-\alpha}^2}{2}$$

where $\hat{\sigma}^2$ = estimated total variance,
z_p = the pth percentile of the standard normal distribution,
α = false rejection decision error rate,
β = false acceptance decision error rate,
Δ = the width of the gray region, and
n = the number of samples.

Select the Sample Size That Satisfies the DQOs for Each Data Collection Design

Simple Random Sampling for Initial Data Quality Objectives

The RDT used the DEFT software to calculate the number of samples needed to meet the initial false rejection and false acceptance error limits specified in Step 6. The DEFT inputs and outputs for this sampling design are summarized in Table D-5. Given the decision error limits of 1 percent, and a gray region from 500-600 mg/kg, the RDT could not afford to implement a simple random sampling design for 90 EUs. This design would have cost over $3.16 million to implement and would have required that 200 samples be collected from each EU. The Decision Performance Goal Diagram generated by DEFT with the inputs for this design option is presented in Figure D-3. Since this design did not even come close to the $50,000 budget, the RDT decided to explore the idea of composite sampling.

Composite Sampling for Initial Data Quality Objectives

DEFT was used to derive the required number of composite samples per DU. The RDT developed DEFT inputs for a composite sampling design in which eight “scoops” were collected from each EU and combined for analysis. Eight “scoops” was considered optimal by the RDT as the field sampling crew was experienced with collecting units of eight, the QC criteria for eight “scoops” clearly described, and from background evidence on the characteristics of the Blue Mountain site, sufficient to be deemed enough to be representative of the are from which they

Table D-5. DEFT Inputs and Outputs for SRS and Composite Sampling for the Initial and Relaxed Error Limits for 90 EUs

Parameter	Initial SRS Input	Initial Composite Input	Relaxed SRS Input	Relaxed Composite Input
Sampling Cost	$100 ea.	$40 ea. (per "scoop")	$100 ea.	$40 ea. (per "scoop")
Analytical Cost	$75 ea.	$75 ea.	$75 ea.	$75 ea.
Action Level	600 mg/kg	600 mg/kg	600 mg/kg	600 mg/kg
Gray Region	500-600 mg/kg	500-600 mg/kg	400-600 mg/kg	400-600 mg/kg
Null Hypothesis (H_0)	Mean [As] ≥ 600 mg/kg	Mean [As] ≥ 600 mg/kg	Mean [As] ≥ 600 mg/kg	Mean [As] ≥ 600 mg/kg
False Acceptance Decision Error	500 mg/kg (p = .01)	500 mg/kg (p = .01)	400 mg/kg (p = .30)	400 mg/kg (p = .30)
False Rejection Decision Error	600 mg/kg (p = .01)	600 mg/kg (p = .01)	600 mg/kg (p = .05)	600 mg/kg (p = .05)
Standard Deviation	302 mg/kg	302 mg/kg	302 mg/kg	302 mg/kg
Number of "scoops"	N/A	8	N/A	8
Measurement SD/	N/A	0.099	N/A	0.099
Parameter	**Initial SRS Output**	**Initial Composite Output**	**Relaxed SRS Output**	**Relaxed Composite Output**
Number samples/EU	201	30	13	3
Cost/EU	$35,175	$11,850	$2,275	$1,185
Total Cost	$3,166,750	$1,066,500	$204,750	$106,650

were drawn. The sampling cost for composite sampling was $40/scoop to reflect the lower cost of collecting, bagging, labeling, and handling of scoops composited into a single sample the field. That means that the total cost of collecting one composite sample was $320 (i.e., $40 x 8). This one composite sample would then be analyzed in the laboratory for $75, so that the total cost of obtaining arsenic concentrations averaged over 8 locations would be $395 (i.e., $320 + $75). For comparison, the cost of obtaining information from 8 locations individually through the simple random sampling design would be $1,400 [i.e., ($100 x 8) + ($75 x 8)].

DEFT calculated that this design would have required that 30 samples be collected from each EU, for a total cost of over $1 million. The Decision Performance Goal Diagram generated by DEFT for this composite sampling design is identical to the one generated for the simple random sampling design, which is presented in Figure D-3. The DEFT inputs and outputs for this design are listed in Table D-5. Since this composite sampling design far exceeded the given

budget of $50,000, the RDT decided to return to Step 6, relax the decision error limits, and expand the size of the gray region.

Simple Random Sampling with Relaxed Error Limits

The RDT discussed the potential consequences of both types of error again. They determined that, since they were not as concerned with making false acceptance errors (wrongly remediating a "clean" EU), that they would relax this limit as far as the representatives of the PRP would allow, which was 30 percent (p = .3). They also agreed that, given the current and suspected future use of the site, the false rejection error rate could be relaxed to 5 percent (p = .05). In addition, they decided to enlarge the gray region from 500-600 mg/kg to 400-600 mg/kg. The DEFT inputs and outputs for this design are summarized in Table D-5. This design would have required that the RDT collect 13 samples per EU, for a total cost of $204,750. The Decision Performance Goal Diagram generated by DEFT with the inputs for this design is presented in Figure D-4. Since the cost of implementing this design was still about four times the budget, the RDT decided to combine composite sampling with these relaxed decision error limits.

Composite Sampling with Relaxed Error Limits

Using the eight “scoop” model and reduced cost of sample collection ($40/scoop), the RDT ran DEFT with the relaxed decision error limits and a wider gray region. The DEFT inputs and outputs for this design are summarized in Table D-5. The Decision Performance Goal Diagram generated by DEFT for these inputs was identical to the diagram generated for the “relaxed” simple random sampling design shown in Figure D-4. Although composite sampling with relaxed error limits cut the total cost to $106,650 but still twice the allocated budget.

The RDT realized that they would have to find another means of generating an appropriate design while remaining within budget. To do this, they turned back to Step 4 of the DQO Process, Define the Boundaries of the Study.

Revisiting Step 4: Simple Random Sampling for Larger Decision Units

The RDT recognized that one of the drivers of cost was the large number of EUs because the sample sizes calculated based on the DQOs had to be applied to each of the 90 EUs. The RDT decided to re-examine the scale of decision making, which was discussed in Step 4, Define the Boundaries. After discussion of typical activities at this site, the risk assessor agreed that ½-acre EUs might be overly conservative, and workers would probably integrate their exposure over much larger areas over a 30-year period. The RDT, therefore, considered partitioning the site into larger Decision Units (DUs).

The RDT determined that they could divide the surface soil OU into four distinct areas based upon the potential threat that the area posed to site workers. The primary surface soils about which they were concerned were those that were commonly traversed by the workers.

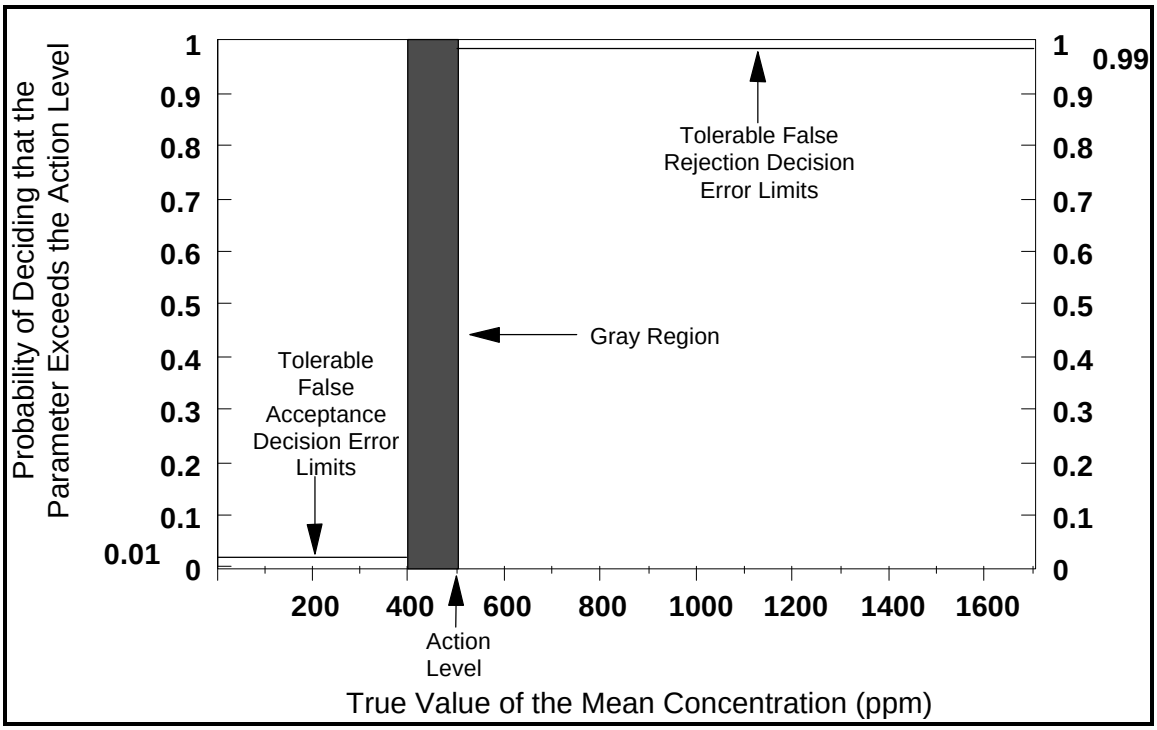


Figure D-3. Decision Performance Goal Diagram for Initial DEFT Inputs

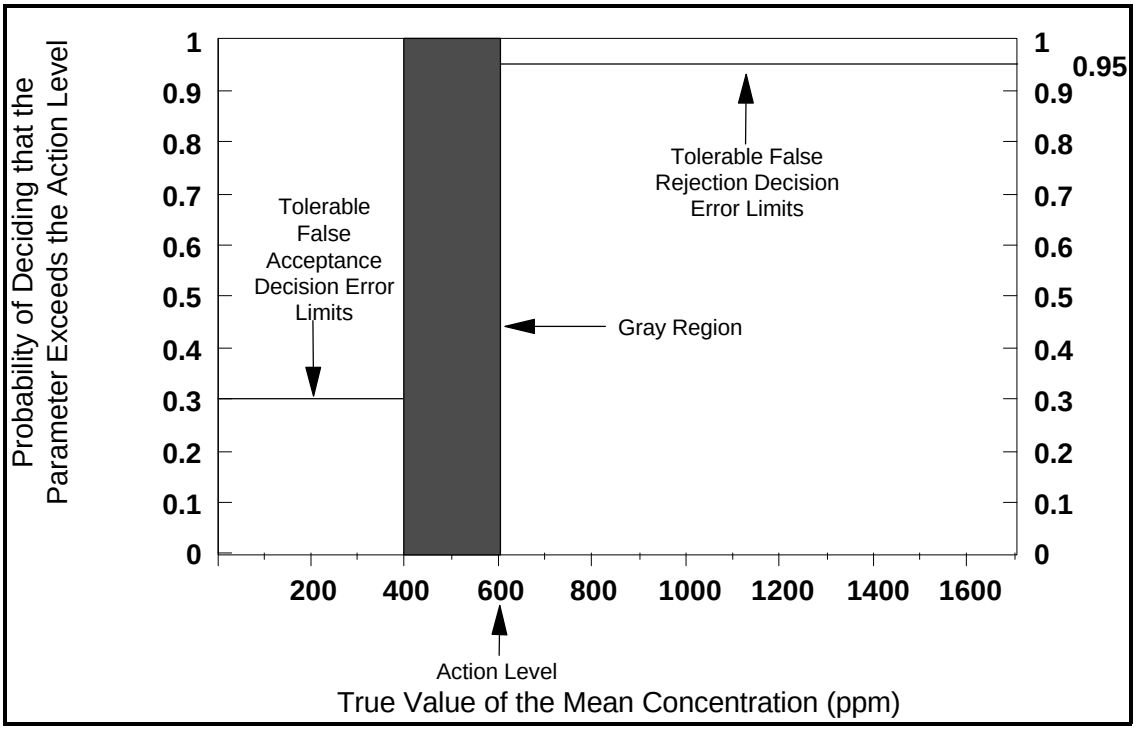


Figure D-4. Decision Performance Goal for Relaxed DEFT Inputs

Figure D-5. Blue Mountain Smelter Site Decision Units

They were also concerned about the surface soils near the slag piles and near the ponds, since these areas were the sites of daily worker activities and would soon become the sites of remedial activities (in some places, remediation had already begun). They were least concerned about remote portions of the site, where workers rarely ventured. The RDT decided to split the site into four distinctly different areas: near-building, near-slag pile, near-pond, and remote (Figure D-5). These four Decision Units were much larger than any of the 90 EUs considered earlier. The approximate surface area of each is listed in Table D-6. Rather than collect data and make decisions for each of the smaller 90 EUs, the team decided to sample and make a decision for each of the four DUs.

The consequences of both types of decision error were far greater for DUs than for EUs, because DUs were so much larger and would result in more wasted resources if sampling falsely indicated that contamination was above the action level, and would pose greater health

Table D-6. DEFT Inputs and Outputs for SRS for Four Decision Units

Parameter	Near-bldg. Input (7 acres)	Near-Slag Pile Input (10 acres)	Near-Pond Input (9 acres)	Remote Input (19 acres)
Sampling Cost	$100 ea.	$100 ea.	$100 ea.	$100 ea.
Analytical Cost	$75 ea.	$75 ea.	$75 ea.	$75 ea.
Action Level	600 mg/kg	600 mg/kg	600 mg/kg	600 mg/kg
Gray Region	500-600 mg/kg	500-600 mg/kg	500-600 mg/kg	500-600 mg/kg
Null Hypothesis (H_0)	Mean [As] ≥ 600 mg/kg	Mean [As] ≥ 600 mg/kg	Mean [As] ≥ 600 mg/kg	Mean [As] ≥ 600 mg/kg
False Acceptance Decision Error Limit	500 mg/kg (p = .20)	500 mg/kg (p = .20)	500 mg/kg (p = .20)	500 mg/kg (p = .20)
False Rejection Decision Error Limit	600 mg/kg (p = .01)	600 mg/kg (p = .05)	600 mg/kg (p = .10)	600 mg/kg (p = .20)
Standard Deviation	302 mg/kg	302 mg/kg	302 mg/kg	302 mg/kg
Parameter	**Near-Bldg. SRS Output**	**Near-Slag Pile SRS Output**	**Near-Pond SRS Output**	**Remote SRS Output**
Number of Samples	95	58	42	27
Cost	$16,625	$10,150	$7,350	$4,725
Total Cost for All Four DUs			**$38,850**	

consequences if sampling falsely indicated that they were below the action level. Recognizing that these larger units carried greater decision error consequences, the RDT revisited Step 6 of the DQO process and produced limits of the decision errors that would apply to the DUs (Table D-6).

The team established a gray region of 500 to 600 mg/kg and a limit of 0.2 was assigned to the false acceptance (deciding that the soil concentration is at least 600 mg/kg when, in fact, it was 300 mg/kg) for three of the four DUs (near-building, near-slag pile, near-pond). A similar limit of 0.2 was assigned to the false acceptance for the remote DU. Because this DU was much larger than the other DUs and more seldomly visited by workers, the financial consequences of making a decision to remediate this DU if it was, in fact, below the 600 mg/kg action level would

be severe. The consequences of false rejection errors, however, depended on the number of persons likely to be exposed if the problem was not addressed. Since most of the workers' daily activities occurred in the vicinity of the buildings, the RDT selected a limit of 0.01 for the acceptable false rejection error rate in the near-building DU. Fewer persons worked near the slag and pond piles, so the RDT selected a limit of 0.05 and 0.1 respectively for these areas. The RDT selected limit of 0.2 for the remote areas, where no workers spent a significant amount of time. These limits are summarized in Table D-6.

DEFT calculated that the total cost of implementing the a simple random sampling design that met the criteria discussed above to be $38,850, which fell well within the $50,000 budget for site sampling and analysis. This design would entail that 95 samples be collected from the near-building DU, 58 samples be collected from the near-slag pile, 42 from the near-pond DUs, and that 27 samples be collected from the remote DU. Although acceptable, the RDT decided to explore the possibility of composite sampling for a DU design.

Composite Sampling for Each Decision Unit

Composite sampling entailed combining soil samples collected within each DU and analyzing all the composited samples. This option reduced the number of samples needed to estimate the mean arsenic level for each DU. The team recognized that some information would be lost if they chose this type of sampling, especially for DUs that tested above the action level. Composite sampling data from DUs that tested positive would not indicate the extent of contaminated surface soil within the DU, whereas simple random sampling data provided more information about contaminant localization within a DU. The RDT noted that simple random sampling data would be useful if data collected in this first round of sampling indicated that the DU needed further investigation, because the RDT could use it to develop a better estimate of variability for second round DQOs. Regardless, the team wanted to explore the possible savings using the compositing approach.

The DEFT inputs and outputs for this sampling design are presented in Table D-7. The total cost for the RDT to implement this design was $16,590, well within the $50,000 budget. The RDT then noted that if the original criteria of the probability of both decision errors being 0.01 had been adhered to, the total cost would have been $47,400 with 30 samples (each containing 8 "scoops") from each DU.

Select The Most Resource-effective Design That Satisfies All The DQOs-In the end, the RDT decided to implement the simple random sampling for four DUs (shown in Table D-6). Although both the SRS and the composite designs for the DU model proved cost-effective, the RDT felt that the simple random sample provided valuable information about contaminant distribution that was lost under the composite design. With a simple random sample, any DU for which the hypothesis test result is negative (failure to reject the idea that the mean equals or exceeds 600 mg/kg) can be easily located and would provide useful information about the extent of contamination for a Phase II investigation. However, if a sample that indicated that arsenic

Table D-7. DEFT Inputs and Outputs for Composite Sampling for Four Decision Units

Parameter	Near-bldg. Input (7 acres)	Near-Slag Pile Input (10 acres)	Near-Pond Input (9 acres)	Remote Input (19 acres)
Sampling Cost	$40 ea. (per scoop)	$40 ea. (per scoop)	$40 ea. (per scoop)	$40 ea. (per scoop)
Analytical Cost	$75 ea.	$75 ea.	$75 ea.	$75 ea.
Action Level	600 mg/kg	600 mg/kg	600 mg/kg	600 mg/kg
Gray Region	500-600 mg/kg	500-600 mg/kg	500-600 mg/kg	500-600 mg/kg
Null Hypothesis (H_0)	Mean [As] ≥ 600 mg/kg	Mean [As] ≥ 600 mg/kg	Mean [As] ≥ 600 mg/kg	Mean [As] ≥ 600 mg/kg
False Acceptance Decision Error Limit	500 mg/kg (p = .20)	500 mg/kg (p = .20)	500 mg/kg (p = .20)	500 mg/kg (p = .10)
False Rejection Decision Error Limit	600 mg/kg (p = .01)	600 mg/kg (p = .05)	600 mg/kg (p = .10)	600 mg/kg (p = .20)
Standard Deviation	302 mg/kg	302 mg/kg	302 mg/kg	302 mg/kg
Number of "scoops"	8	8	8	8
Measurement SD/ Total SD	0.099	0.099	0.099	0.099
Parameter	**Near-Bldg. Composite Output**	**Near-Slag Pile Composite Output**	**Near-Pond Composite Output**	**Remote Composite Output**
Number of Samples	15	9	7	4
Cost	$5,925	$3,555	$2,765	$1,580
Total Cost for All Four DUs			**$13,825**	

contamination was greater than the action limit in a composite sample, the RDT would not know if the high level of contamination was due to a single, highly contaminated "scoop," or if it was due to a number of moderately contaminated "scoops." The RDT decided that this additional information about contaminant variability would provide them with a more complete idea of arsenic contamination in the surface soil and could potentially save enormous future sampling costs.

Conclusions and Results

After considering existing field data and toxicity information, the RDT decided to focus its study on arsenic. A concentration toxicity screen estimated that arsenic contributed approximately 99 percent of the total risk to onsite workers. New data were needed to decide which, if any, of the 90 EUs had unacceptably high levels of arsenic. The RDT utilized the DQO Process to plan a study of arsenic contamination in surface soil. The first pass through the process (the 90-EU scale of decision making) did not end with a satisfactory sampling design. All alternatives exceeded the sampling and analysis budget. A second pass through the process, however, focused on larger DUs and concluded with an affordable design.

Design Alternatives-Decision error limits were established and DEFT software was used to determine the best simple random sampling and composite sampling designs. These designs all had costs that far exceeded the $50,000 that was budgeted for sampling and analysis.

The RDT decided to divide the site into 4 different areas (DUs) and test these rather than the 90 EUs. Decision errors became more critical because the larger areas caused greater consequences of the decision errors. Decision error limits were set for each of the four areas, and DEFT was used to find suitable simple random and compositing designs. Although the compositing design was less costly, the RDT elected to go with the simple random sample plan. If a problem area were found, the data from the simple random sample plan would then be useful in determining the extent and distribution of contamination in that area. Data from composited samples would not serve that purpose.

DQO Outputs-The RDT developed a sampling and analysis plan that:

- Reflected the desired decision performance criteria and the known site situation;
- Provided a basis for project planners to develop a work/QA plan that, if implemented correctly, would produce data of adequate quality and quantity for making the decisions with the desired confidence;
- Developed a sampling design that did not exceed the allocated budget for this effort; and
- Considered future use of the data, ensuring that the data would be helpful in the event that information was needed on the distribution and extent of contamination within decision units.

Statistical Assumptions-The formula used to determine sample size (equation 1, Step 7: Optimize the Design for Data Collection) makes several important assumptions: Normality, Independence, and Estimated Total Variability. The first two assumptions are reasonably impervious to minor failings and do not greatly affect the sample size; the Estimated Total Variability, however, directly affects the sample size. The more precisely this can be estimated or controlled, the lower the number of samples required. In this case, the RDT estimated the field variability to be an order of magnitude greater than the laboratory variability as defined by the standard deviation (standard deviation field = 10 times the standard deviation laboratory) leading to a total standard deviation of 302. If the RDT had defined it differently with respect to variance (variance field = 10 times variance laboratory) it would have led to a total standard deviation of 99.5 (variance laboratory = 30^2, therefore variance field = 10 X 30^2, add together, then the square root taken) and a much lower number of samples required.

This potential source of confusion can only be resolved by estimating the total variance through preliminary sampling. In this case the RDT elected to use the most conservative method to estimate total variance as preliminary data was unavailable. Using a large total variance when it should be much smaller results in a diminuation of potential false rejection and false acceptance decision error rates in decision-making. Properly estimating the total variance enables the gray region to be reduced to a minimum thereby improving decision making.

Results-The RDT completed its sampling design by selecting random sample locations within each DU. They then arranged for collection and analysis of the samples. The data were assessed using the DQA process and only one sample was found to have an arsenic concentration significantly above the action level. This sample was located in the area east of the former ore storage building, indicating the possibility of some localized contamination (a hot spot). Since none of the other samples were significantly above the action level, the RDT turned its focus for the second phase of their investigation to the development of DQOs for surface soils in the “near-building” DU, adjacent to the former ore storage building. The remainder of the site surface soil was characterized as Site Evaluation Accomplished, as indicated by the decision rule of Step 5.