AUG 2000

LANL General 2000

United States
Environmental Protection
Agency

Office of Environmental
Information
Washington, DC 20460

EPA/600/R-96/055
August 2000

EPA

Guidance for the Data Quality Objectives Process

EPA QA/G-4

LIBRARY COPY

Quality

13425

Foreword

The U.S. Environmental Protection Agency (EPA) has developed the Data Quality Objectives Process as the Agency's recommended planning process when environmental data are used to select between two opposing conditions. The Data Quality Objectives Process is used to develop Data Quality Objectives 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. When this Process is not immediately applicable (i.e., the objective of the program is estimation, research, or any other objective that does not select between two distinct conditions), the Agency requires the use of a systematic planning method for defining performance criteria. This document, *Guidance for the Data Quality Objectives Process (EPA QA/G-4)* provides a standard working tool for project managers and planners to develop Data Quality Objectives for determining the type, quantity, and quality of data needed to reach defensible decisions.

As required by EPA Manual 5360 (May 2000), this document is valid for a period of up to five years from the official date of publication. After five years, this document will be reissued without change, revised, or withdrawn from the EPA Quality System series documentation.

This document is one of the *EPA Quality System Series* documents which describe EPA policies and procedures for planning, implementing, and assessing the effectiveness of a quality system. Questions regarding this document or other *EPA Quality System Series* documents should be directed to:

U.S. EPA
Quality Staff (2811R)
1200 Pennsylvania Ave, NW
Washington, DC 20460
Phone: (202) 564-6830
Fax: (202) 565-2441
e-mail: quality@epa.gov

Copies of *EPA Quality System Series* documents may be obtained from the Quality Staff or by downloading them from the Quality Staff Home Page:

www.epa.gov/quality

TABLE OF CONTENTS

		Page
CHAPTER 0.	**INTRODUCTION**	**0-1**
0.1	EPA Program Requirements for Environmental Data and Technology	0-1
0.2	Systematic Planning and the Data Quality Objectives Process	0-4
0.3	Benefits of Using the DQO Process	0-9
0.4	Organization of this Document	0-12
0.5	Background for the Three Examples	0-13
CHAPTER 1.	**STEP 1. STATE THE PROBLEM**	**1-1**
1.1	Background	1-1
1.2	Activities	1-2
1.3	Outputs	1-5
1.4	Examples	1-5
CHAPTER 2.	**STEP 2. IDENTIFY THE DECISION**	**2-1**
2.1	Background	2-1
2.2	Activities	2-2
2.3	Outputs	2-4
2.4	Examples	2-4
CHAPTER 3.	**STEP 3. IDENTIFY THE INPUTS TO THE DECISION**	**3-1**
3.1	Background	3-1
3.2	Activities	3-2
3.3	Outputs	3-4
3.4	Examples	3-4
CHAPTER 4.	**STEP 4. DEFINE THE BOUNDARIES OF THE STUDY**	**4-1**
4.1	Background	4-1
4.2	Activities	4-3
4.3	Outputs	4-7
4.4	Examples	4-8
CHAPTER 5.	**STEP 5. DEVELOP A DECISION RULE**	**5-1**
5.1	Background	5-1
5.2	Activities	5-2
5.3	Outputs	5-4
5.4	Examples	5-5

TABLE OF CONTENTS (continued)

		Page
CHAPTER 6.	**STEP 6. SPECIFY TOLERABLE LIMITS ON DECISION ERRORS**	**6-1**
6.1	Background	6-1
6.2	Activities	6-2
6.3	Outputs	6-12
6.4	Examples	6-14
CHAPTER 7.	**OPTIMIZE THE DESIGN FOR OBTAINING DATA**	**7-1**
7.1	Background	7-1
7.2	Activities	7-2
7.3	Outputs	7-6
7.4	Examples	7-6
CHAPTER 8.	**BEYOND THE DATA QUALITY OBJECTIVES PROCESS**	**8-1**
8.1	Planning	8-1
8.2	Implementation	8-3
8.3	Assessment	8-3
APPENDIX A.	**DERIVATION OF SAMPLE SIZE FORMULA FOR TESTING MEAN OF NORMAL DISTRIBUTION VERSUS AN ACTION LEVEL**	**A-1**
APPENDIX B.	**BIBLIOGRAPHY AND REFERENCES**	**B-1**
APPENDIX C.	**DQO EXAMPLE: USE OF THE MEDIAN**	**C-1**
APPENDIX D.	**GLOSSARY OF TERMS**	**D-1**

LIST OF FIGURES

Page

Figure 0-1. The EPA Quality System 0-2

Figure 0-2. The Scientific Method 0-4

Figure 0-3. The Data Quality Objectives Process 0-6

Figure 0-4. Repeated Application of the DQO Process Throughout the Life Cycle of a Project 0-8

Figure 2-1. An Example of the DQO Process Applied to Multiple Decisions for a Hazardous Site Investigation 2-3

Figure 4-1. Geographic Boundaries Delineated Using a Map 4-4

Figure 6-1. An Example of How Total Study Error Can Be Broken Down by Components. 6-3

Figure 6-2. An Example of a Decision Performance Curve 6-7

Figure 6-3. An Example of a Decision Performance Goal Diagram (Baseline Condition: Parameter exceeds the Action Level) 6-9

Figure 6-4. An Example of a Decision Performance Goal Diagram (Baseline Condition: Parameter is less than the Action Level) 6-13

Figure 6-5. Decision Performance Goal Diagram for Example 1 6-13

Figure 6-6. Decision Performance Goal Diagram for Example 2 6-15

Figure 8-1. The Data Quality Assessment Process 8-4

Figure C-1. Decision Performance Goal Diagram for Dust Lead Loading C-4

LIST OF TABLES

Page

Table 0-1. Elements of the Systematic Planning Process 0-5

Table 2-1. An Example of a Principal Study Question and Alternative Actions 2-2

Table 5-1. Population Parameters and Their Applicability to a Decision Rule 5-3

Table 6-1. False Acceptance and False Rejection Decisions 6-5

Table 7-1. False Acceptance Decision Error Rates and Alternative Sampling Frequencies 7-8

Table 8-1. Elements of a Quality Assurance Project Plan 8-2

Table C-1. Number of Samples Required for Determining if the True Median Dust Lead Loading is Above the Standard C-5

CHAPTER 0

INTRODUCTION

After reading this chapter you should understand the structure and function of EPA's Quality System, the kinds of programs that are a part of this System, and the benefits of using the Data Quality Objectives Process.

When data are being used to select between two alternative conditions (e.g., compliance or non-compliance with a standard), the Agency's recommended systematic planning tool is the Data Quality Objectives (DQO) Process. The DQO Process is a systematic planning process that is part of the EPA's Quality System.

Who can use this guidance document? This guidance is intended for project managers, technical staff, regulators, stakeholders, and others who wish to use the DQO Process to plan data collection efforts and develop an appropriate data collection design to support decision making.

0.1 EPA Quality System Requirements

EPA Order 5360.1 A2 (EPA 2000a) and the applicable Federal regulations establish a mandatory Quality System that applies to all EPA organizations and organizations funded by EPA. Components of the Quality System are presented in Figure 0-1. Organizations must ensure that data collected for the characterization of environmental processes and conditions are of the appropriate type and quality for their intended use and that environmental technologies are designed, constructed, and operated according to defined expectations. Systematic planning is a key project-level component of the EPA Quality System (see Figure 0-1).

EPA policy is based on the national consensus standard, ANSI/ASQC E4-1994, *Specifications and Guidelines for Environmental Data Collection and Environmental Technology Programs,* developed by the American National Standards Institute and the American Society for Quality. This document describes the necessary management and technical area elements for developing and implementing a quality system by using a tiered approach to a quality system. The standard recommends first documenting each organization-wide quality system in a Quality Management Plan or Quality Manual (to address requirements of *Part A: Management Systems* of the standard), and then documenting the applicability of the quality system to technical activity-specific efforts in a Quality Assurance Project Plan or similar document (to address the requirements of *Part B: Collection and Evaluation of Environmental Data* of the standard). EPA has adopted this tiered approach for its mandatory Agency-wide Quality System. This document addresses Part B requirements of the standard for systematic planning for environmental data operations.

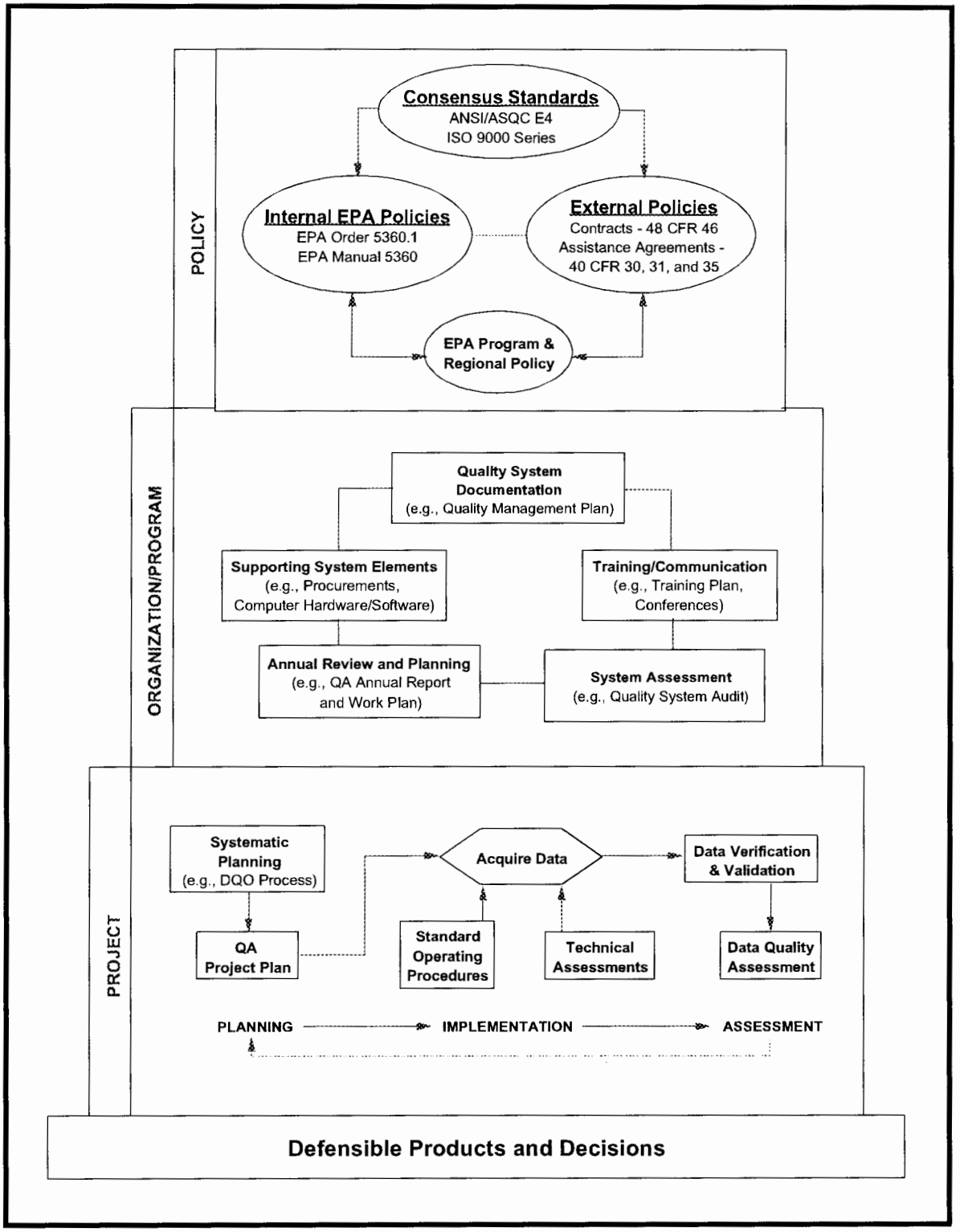


Figure 0-1. EPA Quality System Components and Tools

In accordance with EPA Order 5360.1 A2, the Agency requires that:

- Environmental programs performed for or by the Agency be supported by data of the type and quality appropriate to their expected use. EPA defines environmental data as information collected directly from measurements, produced from models, or compiled from other sources such as databases or literature.

- Decisions involving the design, construction, and operation of environmental technology be supported by appropriate quality assured engineering standards and practices. Environmental technology includes treatment systems, pollution control systems and devices, waste remediation, and storage methods.

EPA Order 5360.1 A2 is supported by the *EPA Quality Manual for Environmental Programs* (U.S. EPA, 2000b) that defines requirements for implementing EPA's Quality System. The Order defines the quality requirements and the Manual presents the mandatory "how to" for implementing these requirements.

EPA's Quality System (presented in Figure 0-1) comprises three levels – Policy, Organization/Program, and Project:

- Policy – this level addresses Agency-wide quality policies and regulations that both EPA organizations and external EPA-funded organizations must address;

- Organization/Program – this level addresses the management and implementation component of the individual Quality System; and

- Project – this level addresses the project-specific components that are applied to individual projects to ensure that the needs of the organization are met.

EPA has developed a *Quality System Series* of documents that provide guidelines to help organizations ensure that data collected for the characterization of environmental processes and conditions are of the appropriate type and quality for their intended use. Documents useful in planning for data collection include:

- *Decision Error Feasibility Trials (DEFT) Software for the Data Quality Objectives Process (EPA QA/G-4D)*,
- *Guidance for the Data Quality Objectives Process for Hazardous Waste Sites (EPA QA/G-4HW)*,
- *Guidance on Quality Assurance Project Plans (EPA QA/G-5)*,
- *Guidance for the Preparation of Standard Operating Procedures for Quality-Related Documents (EPA QA/G-6)*, and
- *Guidance for Data Quality Assessment: Practical Methods for Data Analysis (EPA QA/G-9)*.

0.2 Systematic Planning and the DQO Process

EPA Order 5360.1 A2 requires that all EPA organizations (and organizations with extramural agreements with EPA) follow a systematic planning process to develop acceptance or performance criteria for the collection, evaluation, or use of environmental data. A systematic planning process is the first component in the *planning phase* of the project tier, while the actual data collection activities are in the *implementation phase* of this tier (Figure 0-1).

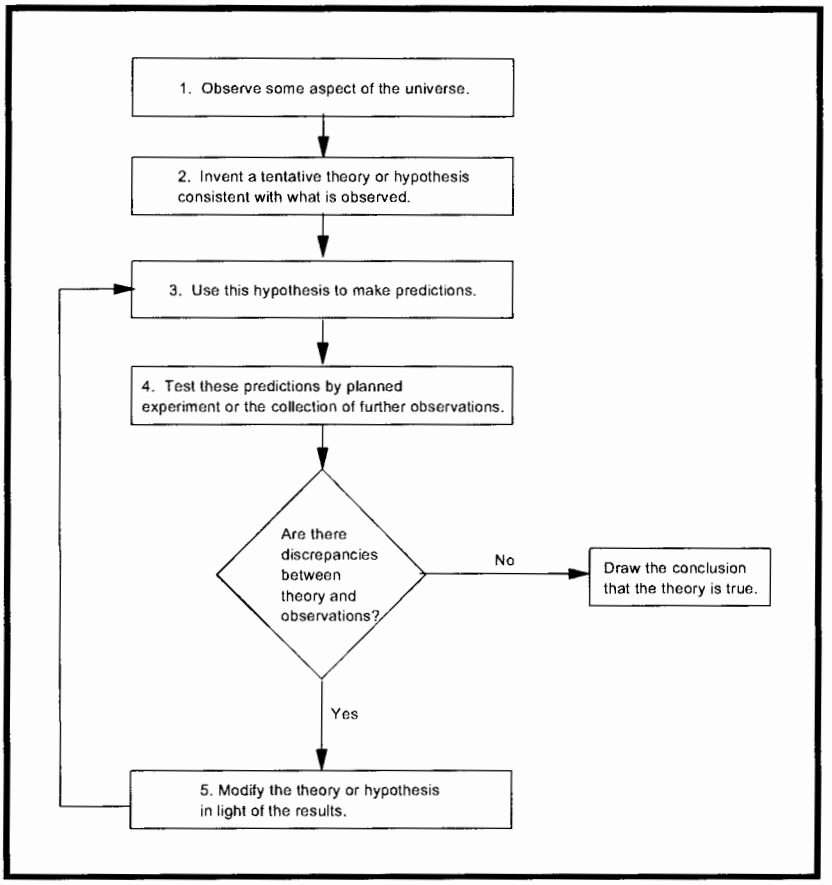


Figure 0-2. The Scientific Method

What is systematic planning? Systematic planning is a planning process that is based on the scientific method and includes concepts such as objectivity of approach and acceptability of results (Figure 0-2). Systematic planning is based on 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. This framework promotes communication between all organizations and individuals involved in an environmental program. Through a systematic planning process, a team can develop acceptance or performance criteria for the quality of the data collected and for the quality of the decision. When these data are being used in decision

making by selecting between two clear alternative conditions (e.g., compliance/non-compliance with a standard), the Agency's recommended systematic planning tool is called the DQO Process. Elements of the systematic planning process (from Section 3.3.8 of the EPA Quality Manual) and relationship to the DQO Process are shown in Table 0-1.

Table 0-1. Elements of the Systematic Planning Process

Elements of Systematic Planning Process	Corresponding Step in the DQO Process
Identifying and involving the project manager/decision maker, and project personnel	Step 1. Define the problem
Identifying the project schedule, resources, milestones, and requirements	Step 1. Define the problem
Describing the project goal(s) and objective(s)	Step 2. Identify the problem
Identifying the type of data needed	Step 3. Identify information needed for the decision
Identifying constraints to data collection	Step 4. Define the boundaries of the study
Determining the quality of the data needed	Step 5. Develop a decision rule Step 6. Specify limits on decision errors
Determining the quantity of the data needed	Step 7. Optimize the design for obtaining data
Describing how, when, and where the data will be obtained	Step 7. Optimize the design for obtaining data
Specifying quality assurance and quality control activities to assess the quality performance criteria	Part B of QA Project Plan
Describing methods for data analysis, evaluation, and assessment against the intended use of the data and the quality performance criteria	Part D of QA Project Plan; DQA Process

What are acceptance or performance criteria? Acceptance or performance criteria are based on the ultimate use of the data to be collected and needed quality assurance (QA) and quality control (QC) practices required to support the decision. In the decision making process, these criteria allow a user to limit decision errors to a fixed level for determining whether or not an Action Level (regulatory or risk-based) has been exceeded.

What is the DQO Process? The DQO Process is a seven-step planning approach to develop sampling designs for data collection activities that support decision making. This process uses systematic planning and statistical hypothesis testing to differentiate between two or more clearly defined alternatives. A summary of the seven steps is presented in Figure 0-3.

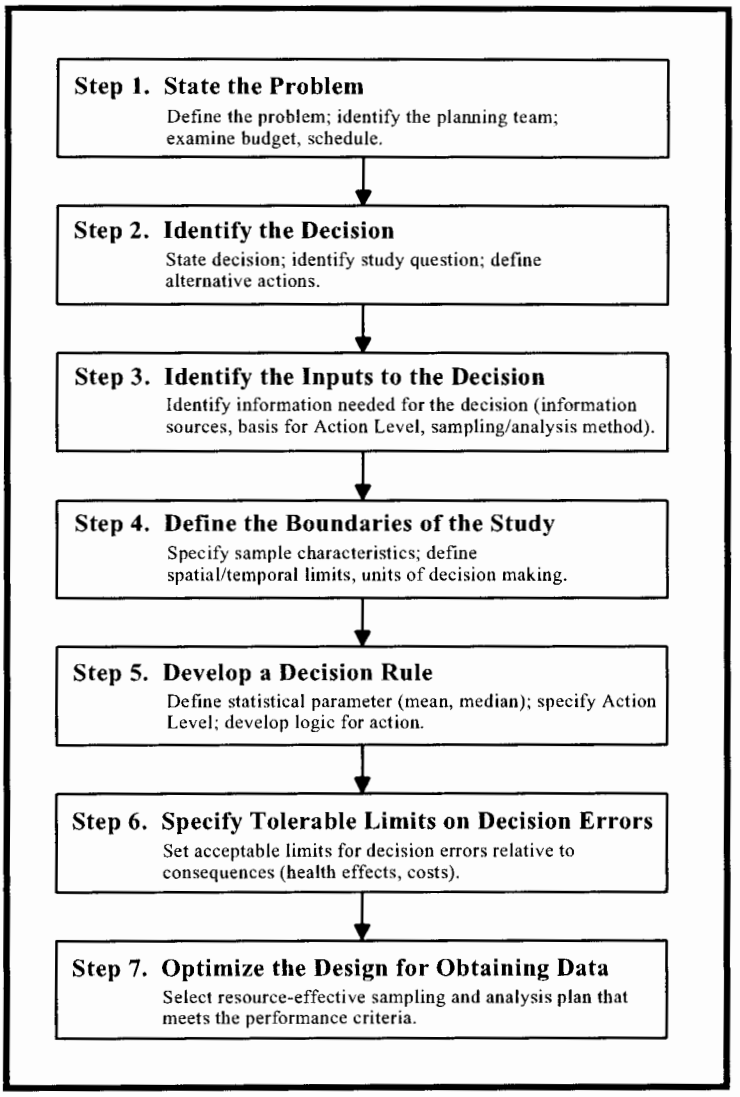


Figure 0-3. The Data Quality Objectives Process

The DQO Process is iterative and allows the planning team to incorporate new information and modify outputs from previous steps as inputs for a subsequent step. Although the principles of systematic planning and the DQO Process are applicable to all scientific studies, the DQO Process is particularly designed to address problems that require making a decision between two clear alternatives. The final outcome of the DQO Process is a design for collecting data (e.g., the number of samples to collect, and when, where, and how to collect samples), together with limits on the probabilities of making decision errors.

What are DQOs? DQOs are qualitative and quantitative statements, developed using 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. DQOs define the performance criteria that limit the probabilities of making decision errors by considering the purpose of collecting the data; defining the appropriate type of data needed; and specifying tolerable probabilities of making decision errors.

What projects are covered by the DQO Process? The DQO Process may be applied to all programs involving the collection of environmental data used in decision making. The principles used in the DQO Process are also applicable to programs with objectives other than decision making (e.g., estimation and research studies).

Who should be included in the DQO Process? When applying the DQO Process, a planning team of senior program staff, technical experts, managers, data users (usually with some statistical expertise), a quality assurance specialist, regulators, and stakeholders are usually involved. It is important that the key persons participate (or stay informed) throughout the DQO Process so that each individual understands the problem/decision and objectives of the decision-making process. Individuals with specific areas of technical expertise may decide to be involved only in the steps of the DQO Process that require technical input.

When should the DQO Process be used? The DQO Process should be used during the planning stage of any study that requires data collection, *before* the data are collected. As the DQO Process is iterative by nature, steps within the process can be revisited before a final decision is reached. As shown in Figure 0-4, the planning team may choose to revisit selected parts of the DQO Process or to investigate the entire process cyclically.

Is the DQO Process only applicable to large studies or studies that require multiple decisions? The DQO Process applies to any study, regardless of its size. However, the depth and detail of DQO development will depend on the study objectives. The DQO Process is particularly applicable to a study in which multiple decisions must be reached because, by using this planning process, the planning team can clearly separate and delineate data requirements for each problem/decision. For projects that require multiple decisions or answers to more than one question, it is likely that the resolution of one decision will lead to the evaluation of subsequent decisions. In these cases, the DQO Process can be used repeatedly throughout the life cycle of a project. Often, the decisions that are made early in the project will be preliminary in nature; they might require only a limited planning and evaluation effort. As the study nears conclusion and the consequences of making a decision error become more critical, however, the level of effort needed to resolve a decision generally will become greater. Figure 0-4 illustrates this point.

What are the outputs of the DQO Process? The DQO Process leads to the development of acceptance or performance criteria based on the ultimate use of the data to be collected and define the quality required for the decision in terms of acceptance limits on the probabilities of committing a decision error. Each step of the DQO Process defines criteria that will be used to

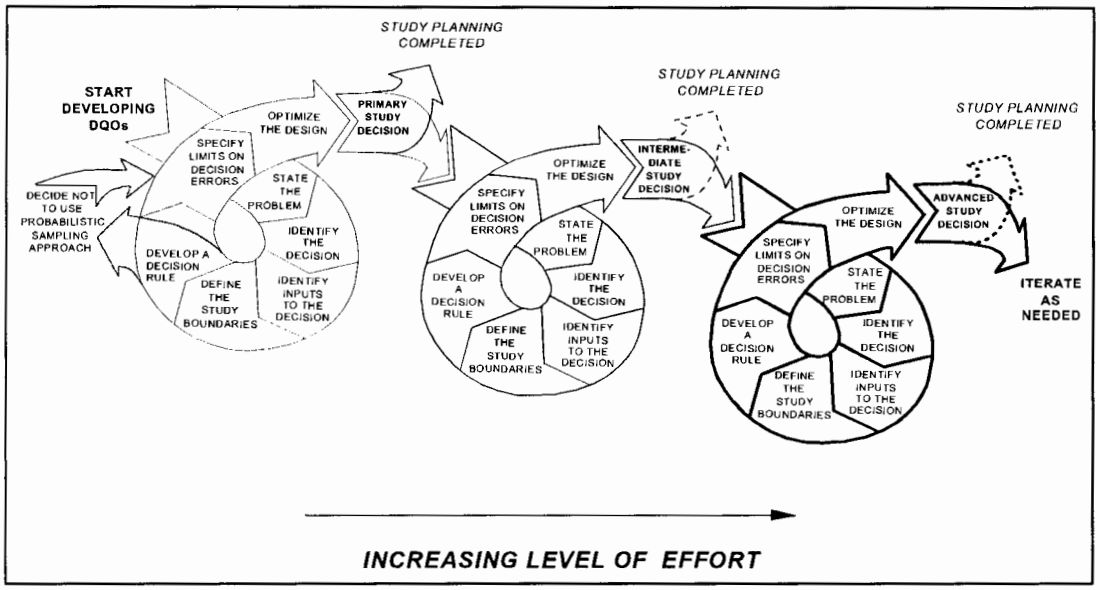


Figure 0-4. Repeated Application of the DQO Process throughout the Life Cycle of a Project

establish the final data collection design. The first five steps of the DQO Process are primarily focused on identifying qualitative criteria, such as:

- the nature of the problem that has initiated the study and a conceptual model of the environmental hazard to be investigated;
- the decisions that need to be made and the order of priority for resolving them;
- the type of data needed (i.e., geographic area, environmental medium, overall timing of data collection, etc.); and
- a decision rule that defines how the data will be used to choose among alternative actions.

The sixth step defines quantitative criteria, expressed as limits on the probability or chance (risk) of making a decision error, that the decision maker can tolerate. The seventh step is used to develop a data collection design based on the criteria developed in the first six steps. In this step the planning team considers the final product of the DQO Process, a data collection design that meets the quantitative and qualitative needs of the study using a specified number of samples that can be accommodated by the budget available. The outputs of the DQO Process are used to develop a QA Project Plan and for performing Data Quality Assessment (Chapter 8).

What is a data collection design? A data collection design specifies the number, location, physical quantity, and type of samples that should be collected to satisfy the DQOs. The sampling design designates where, when, and under what conditions samples should be collected; what variables are to be measured; and the QA and QC activities that will ensure that sampling design

and measurement errors are managed sufficiently to meet the tolerable decision error rates specified in the DQOs. These QA and QC activities together with details of the data collection design are documented in the QA Project Plan.

Can existing data be used in the DQO Process to support your decision making? Existing data can be very useful. For example, pilot studies are often performed to provide a preliminary assessment of variability. In these cases, the existing data may provide valuable information to help develop a design for collecting data. It is critical to examine the existing data to ensure that their quality is acceptable for use, or for integration into a new data set. Some considerations include:

- determining if the existing data were collected within approximately the same spatial and temporal boundaries as the new data;
- examining the existing data to determine if this data set includes identical media and analytes;
- examining the performance of the analytical methods for the existing data (accuracy, precision, detection limits) and comparing this to the specifications in Step 3 of the DQO Process for new data to be collected; and
- examining the variability among samples in the existing and new data sets.

Combining existing data and new data can be a very complex operation and you should undertake this with great care. In many cases, statistical expertise is required to evaluate both data sets before they can be combined with confidence.

Will you always develop statistical/probabilistic sampling designs for data collection if you use the DQO Process? No. Although statistical methods for developing the data collection design are strongly encouraged, this guidance recognizes that not every sampling problem can be resolved with probabilistic sampling designs. However, the DQO Process can and should be used as a planning tool for studies even if a statistical data collection design ultimately will not be used. In these cases, the planning team is encouraged to seek expert advice on how to develop a non-statistical data collection design and how to evaluate the results of the data collection. When nonprobabilistic, judgmental, or quota sampling methods are used, be sure to consult with an EPA representative to ensure that program-specific QA requirements are satisfied.

How should you use this guidance? You should use this guidance as a tool to structure the planning activities for collecting environmental data. It should be used to organize meetings, focus the collection of background information, and facilitate communication between a team that includes technical experts, program managers, stakeholders, regulators, and decision makers.

0.3 Benefits of Using the DQO Process

The DQO Process integrates a multidisciplinary team and offers the advantages of using experience and resources of individuals who have different backgrounds, different kinds of

knowledge, and who can collectively focus on achieving a successful project conclusion. During the initial planning stages, the planning team can concentrate on developing requirements for collecting the data and work to reach consensus on the type, quantity, and quality of data needed to support Agency decisions. This interaction results in a clear understanding of the problem and the options available for addressing it, the development of acceptance or performance criteria for decision making, a consensus-based approach to understanding the problem, and data being collected of appropriate quality. Organizations that have used the DQO Process have observed that:

- The structure of the DQO Process provides a convenient way to *document activities and decisions* and to communicate the data collection design to others. This documentation *facilitates rapid review and approval* by regulators and stakeholders.

- The DQO Process enables data users and relevant technical experts to participate collectively in data collection planning and to specify their particular needs prior to data collection. The DQO process fosters *communication among all participants*, one of the central tenets of quality management practices, and directs efforts to *achieving consensus* between decision makers, stakeholders, and regulators.

- The DQO Process helps to focus studies by encouraging data users to *clarify vague objectives* and to limit the number of decisions that will be made. Due to this clarification, *the consequences of decision errors are examined* and correct decisions will be made most frequently when the DQO Process is employed.

- The DQO Process is a planning tool that can *save resources* by making data collection operations more resource-effective. Good planning will streamline the study process and increase the likelihood of efficiently collecting appropriate and useful data.

- The DQO Process provides a method for *defining decision performance requirements* that are appropriate for the intended use of the data. This is done by considering the consequences of decision errors and then placing tolerable limits on the chance that the data will mislead the decision maker into committing a decision error. A statistical sampling design can then be generated to provide the most efficient method for managing decision errors and satisfying the DQOs.

Upon implementing the DQO Process, your environmental programs may be strengthened by:

- focused data requirements and optimized design for data collection,
- use of clearly developed work plans for collecting data in the field,
- uniformly documented data collection, evaluation, and use,
- clearly developed analysis plans,

- sound, comprehensive quality assurance project plans, and
- up-front buy-in by stakeholders to the sampling design and data collection process.

This can lead to:

- rapid review by regulators and other stakeholders,
- defensible results on which to base decisions,
- increased credibility with regulators and stakeholders, and
- a better use of resources.

Where else can the DQO Process be applied? The DQO Process is widely applicable. For example, the Department of Energy Environmental Management program considers the following potential applications for the DQO Process (Grumley, 1994):

- Waste management
  - Characterizing waste, using process knowledge verified by minimal sampling/ analysis data to meet acceptance criteria for treatment, storage, and disposal.
  - Designing optimal monitoring networks for ground water and surface water discharges, and air emissions.
- Environmental restoration
  - Focusing regulatory and public concerns associated with remediation.
  - Identifying target analytes of concern for remedial activities.
  - Determining when remediation has met cleanup levels.
- Facility transition and management
  - Performing characterization assessments, using existing information or collecting new data, to verify facilities for environmental management acceptance.
  - Evaluating alternative end-state conditions and planning facility deactivation in preparation for eventual decontamination and decommissioning.
  - Designing optimized short- and long-term environmental monitoring.
- Decontamination and decommissioning
  - Determining the location and levels of facility contamination.
  - Determining when decontamination and decommissioning is complete.

- Technology development
  - Determining what constitutes and acceptably demonstrates success in technology development and evaluation.

0.4 Organization of This Document

This document provides EPA's guidance specific to the design plans for collecting data for decision-making activities. EPA recognizes that by using systematic planning and the DQO Process to design environmental data collection efforts, the effectiveness, efficiency, and defensibility of decisions will be improved. This document presents:

- the advantages of using systematic planning for data collection,
- the seven steps of the DQO Process, including activities and outputs for each step, and
- three scenarios that each use a different statistical parameter (mean, median, and upper percentile) to develop a design for collecting data.

The objective of this guidance document is to describe how a planning team can use the DQO Process to generate a plan to collect data of appropriate quality and quantity for defensible decision making. This guidance replaces in its entirety EPA's September 1994 document, *Guidance for the Data Quality Objectives Process (EPA QA/G-4),* (U.S. EPA, 1994a), and is consistent with the *Data Quality Objectives Process for Hazardous Waste Site Investigations (EPA QA/G-4HW)* (U.S. EPA, 1999).

This document contains an introductory chapter that is followed by seven chapters that correspond to the seven steps of the DQO Process. Each chapter is divided into four sections:

1. Background — Provides background information on the DQO Process step, including the rationale for the activities in that step and the objective(s) of the chapter.
2. Activities — Describes the activities recommended for completing the DQO Process step, including how inputs to the step are used.
3. Outputs — Identifies the results that may be achieved by completing the DQO Process step.
4. Examples — Presents outputs from two different DQO scenarios for environmental contamination.

Chapter 8 shows how outputs of the DQO Process are used to develop a QA Project Plan.

Appendix A shows the derivation of the formula used to calculate sample size, and Appendix B gives a Bibliography of referenced books, papers, and publications. Appendix C

shows a complete DQO example using the median as the parameter of interest and Appendix D contains a glossary of terms used in this document.

0.5 Background for the Two Examples

The following examples have been derived from real-life DQO development efforts to illustrate the use of mean and percentile in planning for decision making:

Example 1 - Use of the mean to make a decision about waste disposal of material.

Example 2 - Use of the percentile to make a decision relative to a regulatory limit value.

Example 1. Making Decisions About Incinerator Fly Ash for RCRA Waste Disposal

Cadmium is primarily used for corrosion protection on metal parts of cars, electrical appliances, and in some batteries. Cadmium and cadmium salts have been shown to be toxic to humans through both ingestion and inhalation. Ingestion of concentrations as low as 0.1 mg/kg/day causes mild to severe irritation of the gastrointestinal tract. Exposure from chronic (long-term) inhalation can cause increased incidence of emphysema and chronic bronchitis, as well as kidney damage.

A waste incineration facility located in the Midwest routinely removes fly ash from its flue gas scrubber system and disposes of it in a municipal landfill. Previously the waste fly ash was determined not to be hazardous according to RCRA program regulations. The incinerator, however, recently began accepting and treating a new waste stream. The representatives of the incineration company are concerned that the waste fly ash in a new waste stream could contain hazardous levels of cadmium from new waste sources. They have decided to test the ash to determine whether it should be sent to a hazardous waste landfill or continue to be sent to the municipal landfill.

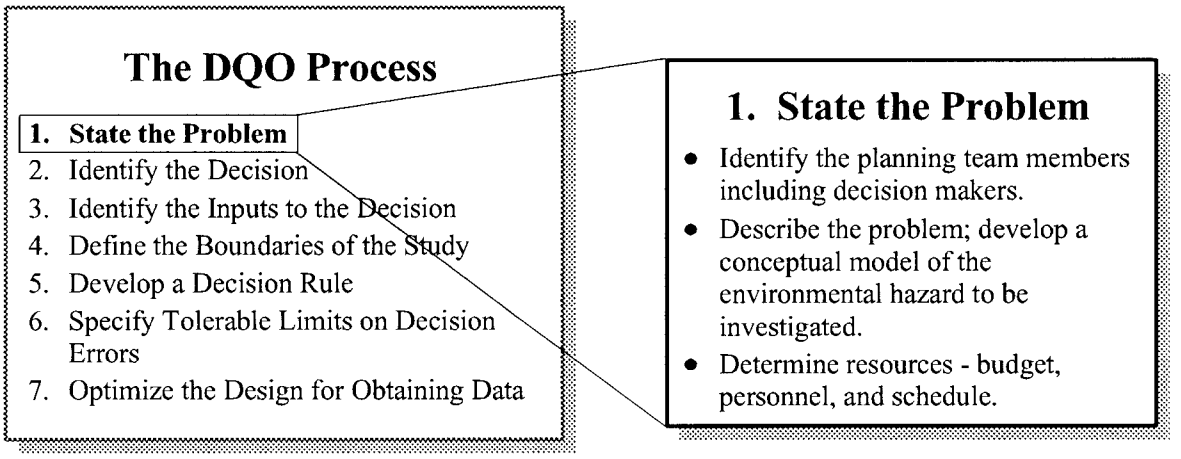
As a precursor to the DQO Process, the incineration company has conducted a pilot study of the fly ash to determine the variability in the concentration of cadmium within loads of waste fly ash leaving the facility and has determined that each load is fairly homogeneous. There is considerable variability between loads, however, due to the nature of the waste stream. The company has decided that testing each container load before it leaves the facility would be an economical approach to evaluating the potential hazard. They could then send containers of ash that exceeded the regulated standards to the higher cost RCRA landfills and continue to send the other containers to the municipal landfill. This example demonstrates use of the mean as the population parameter of concern. (The derivation of a sampling design using the mean is provided in Appendix A).

Example 2. Making Decisions About Urban Air Quality Compliance

In July 1997, the EPA established new ambient air quality standards for $PM_{2.5}$, particulate matter smaller than 2.5 microns (40 CFR 50). $PM_{2.5}$ is comprised of fine particles about 1/30th the thickness of a human hair that are a complex mixture of acids, metals, and carbon. Because the health risks of the chemical components of $PM_{2.5}$ are not fully understood, EPA is implementing $PM_{2.5}$ standards and investigating scientific uncertainties associated with these components.

This example involves monitoring urban air for the presence of $PM_{2.5}$. Representatives of a primary metropolitan statistical area (PMSA) in the northeast wish to determine whether their PMSA is in attainment for $PM_{2.5}$ according to the National Ambient Air Quality Standards (NAAQS). If determined to be in nonattainment, control strategies will be implemented for the PMSA, as defined in its associated State Implementation Plan (SIP). This example uses an upper percentile as the primary population parameter of concern as it is specified in the Standards. Additionally, this example highlights DQO activities and outputs for the case when a data collection design (i.e., number of samples) already has been determined, but not necessarily in accordance with the DQO Process.

CHAPTER 1
STEP 1. STATE THE PROBLEM



After reading this chapter you should understand how to assemble an effective planning team and how to describe the problem and examine your resources for investigating it.

1.1 Background

The first step in any systematic planning process is to define the problem that has initiated the study. As environmental problems are often complex combinations of technical, economic, social, and political issues, it is critical to the success of the process to separate each problem, define it completely, and express it in an uncomplicated format. A proven effective approach to solving a problem is to use a planning team composed of experts and stakeholders who have multidisciplinary backgrounds. A team of individuals with diverse backgrounds offers:

- the ability to develop a complete, concise description of complex problems, and
- multilateral experience and awareness of potential data uses.

If there is a potential that the data collected in the current investigation could be used in future studies (secondary uses of the data), it is important to consult, if possible, potential data users during the planning process.

1.2 Activities

The most important activities in this step are to:

- establish the planning team including the decision makers;
- describe the problem and develop a conceptual model of the environmental hazard to be investigated; and
- identify available resources, constraints, and deadlines.

How do you establish the planning team and decision makers? The DQO planning team is usually composed of the project manager, technical staff, data users (including those with a statistical background), and stakeholders. It is important to carefully select the planning team and leaders (or decision makers) because this team will work together through all seven steps of the planning process. The development of DQOs does not necessarily require a large planning team, particularly if the problem appears to be straightforward. The size of the planning team is usually directly proportional to the complexity and importance of the problem. As the DQO Process is iterative, team members may be added to address areas of expertise not initially considered.

Prior to or during the first meeting of the DQO team, members should identify the decision makers. The decision maker may be one or more individuals familiar with the problem, or with a vested interest in it. As the *technical project manager* is familiar with the problem and the budgetary/time constraints the team is facing, he or she will usually serve as one of the decision makers and will actively participate in all steps of DQO development. The *decision makers* will have the ultimate authority for making final decisions based on the recommendations of the planning team. In cases where the decision makers cannot attend DQO planning meetings, alternate staff members should attend and keep the decision makers informed of important planning issues.

The *technical staff and data users* should include individuals who are knowledgeable about technical issues (such as geographical layout, sampling constraints, analysis, statistics, and data interpretation). The planning team of multidisciplinary experts may include quality assurance managers, chemists, modelers, soil scientists, engineers, geologists, health physicists, risk assessors, field personnel, regulators, and data users with statistical experience.

Stakeholders are individuals or organizations who are directly affected by a decision, interested in a problem, and want to be involved, offer input, or seek information. Usually stakeholders will have multiple perspectives about a problem. The involvement of stakeholders early on in the DQO Process can provide a forum for communication as well as foster trust in the decision making process. An environmental example is the Common Sense Initiative Council, a group of stakeholders convened to offer EPA advice and recommendations on a number of topics. The Common Sense Initiative Council recognizes that involving stakeholders improves communication and assists in analyzing situations to determine the tools and expertise needed to address problems and maintain lasting agreements.

The identification of stakeholders is influenced by the issues under consideration, as well as the ability of stakeholders to articulate their interests. Because EPA is organized into multiple program areas that are concerned with different environmental media that address different regulatory areas (e.g., Clean Air Act, Resource Conservation and Recovery Act), stakeholder involvement activities are not centralized. EPA has developed a web page Introduction to Stakeholder Involvement (http://www.epa.gov/ooaujeag/stakeholders/people.htm) that identifies individuals in various EPA program offices who can offer assistance in stakeholder involvement activities. EPA provides additional information/resources on stakeholder involvement, including:

- EPA Resources for Non-Profit Organizations,
- Children's Health Protection Advisory Committee,
- EPA Voluntary Programs, and
- Partners of Wetlands, Oceans, and Watersheds.

Information for stakeholder involvement and consensus building processes for other federal agencies is also provided at this website. At the state level, information on potential stakeholders is often available. For example, the State of California has developed a directory of citizen groups, government agencies, and environmental education programs concerned with California environmental issues (Harbinger Communications, 1996).

You should identify the roles of team members and group members that have key and secondary roles, then consider the roles of the planning team members when coordinating meetings. While it is important for key members (e.g., decision makers and members involved in policy decisions) to either attend all meetings, or designate a representative to attend meetings that are missed, technical members (e.g., technical managers, field and laboratory personnel, data users, statisticians) may decide to be involved only in meetings where technical input is required. Stakeholders and regulators may elect to attend initial meetings, but miss meetings that address technical issues (e.g., sampling and analysis). When possible, the use of a facilitator or recorder at these meetings is encouraged.

How do you describe the problem and the environmental hazard to be investigated? In Step 1, the planning team describes the conditions or circumstances that are causing the problem and the reasons for undertaking the study. Typical examples for environmental problems include conditions that may pose a threat to human health or the environment or circumstances of potential noncompliance with regulations.

The team may be able to describe the problem as it is currently understood by briefly summarizing existing information, or they may conduct literature searches and examine past or ongoing studies. This will ensure that the problem is correctly defined and has not been solved previously. As you define the problem, you should consider similar studies and document information about the performance of sampling and analytical methods observed in these studies. This information may prove to be particularly valuable later in the DQO Process. You should

organize and review all relevant information, indicate the source of the information, and evaluate its reliability.

The planning team should:

- examine the study objectives from a regulatory standpoint as necessary;
- identify individuals or organizations who are involved or have an interest in the study;
- examine political issues associated with the study;
- look at results of similar studies performed previously from the standpoint of:
  - study parameters,
  - regulatory or other constraints on sampling designs,
  - variability and quality of data collected;
- consider non-technical issues that may influence the sample design; and
- examine possible future uses of the data to be collected (e.g., the data to be collected may be eventually linked to an existing database).

It is **critical** to carefully develop an accurate conceptual model of the environmental problem in this step of the DQO Process, as this model will serve as the basis for all subsequent inputs and decisions. Errors in the development of the conceptual model will be perpetuated throughout the other steps of the DQO Process and are likely to result in developing a sampling and analysis plan that may not achieve the data required to address the relevant issues.

The conceptual model of the potential environmental hazard developed at the beginning of the DQO Process is often a diagram that shows:

- known or expected locations of contaminants,
- potential sources of contaminants,
- media that are contaminated or may become contaminated, and
- exposure scenarios (location of human health or ecological receptors).

If the problem is complex, the team may consider breaking it into more manageable pieces, which might be addressed by separate studies. Priorities may be assigned to individual segments of the problem and the relationship between the segments examined.

How do you identify available resources, constraints, and deadlines? You should examine limitations on resources and time constraints for collecting data. This estimate should include developing acceptance or performance criteria, preparing the QA Project Plan, collecting and analyzing samples, and interpreting data. At this time the planning team should also examine available personnel, and contracts (if applicable) and identify intermediate and final deadlines for collecting data.

1.3 Outputs

The major outputs of this step are:

- a list of the planning team members and their roles,
- identification of decision makers,
- a concise description of the problem and a conceptual model of the environmental problem to be investigated, and
- a summary of available resources and relevant deadlines for the study including budget, availability of personnel, and schedule.

1.4 Examples

Given the background of the three examples as outlined in Section 0.5, the following DQO Step 1 outputs were derived.

Example 1. Making Decisions About Incinerator Fly Ash for RCRA Waste Disposal

How were the planning team members selected? *The planning team included the incineration plant manager, a plant engineer, a quality assurance specialist with statistical experience, and a chemist with sampling experience in the RCRA program. The plant manager was the decision maker.*

How was the problem described and a conceptual model of the potential hazard developed? *The problem was described as determining which container loads of waste fly ash from a new waste stream needed to be sent to a RCRA landfill as a result of a change in an incinerator process that possibly increased the levels of cadmium in waste fly ash. The plant manager wanted to avoid expensive RCRA disposal of waste, if possible, but also needed to comply with regulations and permits.*

The conceptual model described fly ash that was created from industrial waste incineration and is a potential source of toxic metals that include cadmium. Ash is transferred to large containers via a conveyer belt. Containers are filled and trucked to a disposal site. If the waste fly ash is hazardous but disposed in a municipal (sanitary) landfill, then metals can leach into ground water and create runoff to streams, and other surface water bodies, which could pose a hazard to human health and ecological receptors. If such waste is disposed in a RCRA approved landfill, the hazards are contained.

What were the available resources and relevant deadlines? *Although the project was not constrained by cost, the waste generator (the incineration company) wished to hold sampling costs below $2,500. The incineration company also requested that the testing of the waste fly ash in each container be completed within one week.*

Example 2. Making Decisions About Urban Air Quality Compliance

How were the planning team members selected? *The planning team included senior program staff, technical experts, senior managers, a QA specialist, and an individual with statistical experimental design expertise. The most senior program staff member served as the decision maker.*

How was the problem described and a conceptual model of the potential hazard developed? *EPA had set NAAQS for fine particulate matter ($PM_{2.5}$) and other air pollutants (40 CFR 50). The problem was described as determining whether the primary metropolitan statistical area (PMSA) of concern was in attainment for fine particulate matter.*

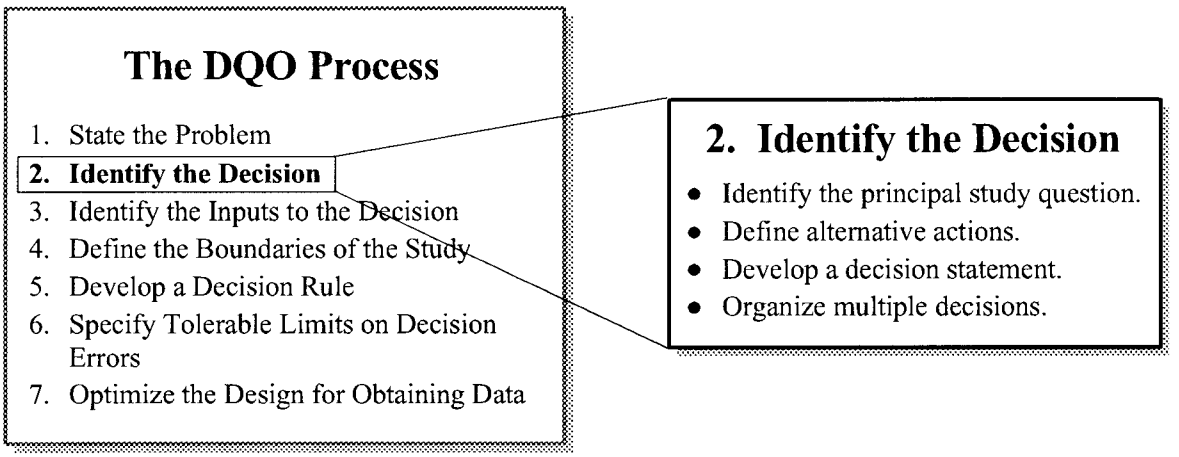
The conceptual model of the potential hazard was considering the concentration of fine particulates in urban air that were primarily combustion products from point and mobile sources. The particulates posed potential sources of exposure from inhalation. As a rule, the PMSA was not concerned with long-term transport because over time particulates aggregated or became deposited on other materials such that the particles came within the purview of the Pm_{10} rule. The PMSA developed a Cartesian map indicating local $PM_{2.5}$ point sources, main roadways, and predominant wind patterns to identify areas of maximum potential exposure.

What were the available resources and relevant deadlines? *The monitoring network was already in place. It consisted of three fixed-site multiple filter gravimetric devices for measuring daily concentrations (24-hr average) once every 3 days. Thus, about 365 readings were obtained each year.*

Looking Ahead to other DQO Steps:

- Careful description of the problem will assist in Step 3, Identify the Inputs to the Decision, when considering additional use of data (link to databases, etc.).
- The conceptual model will assist in Step 4, Define the Boundaries of the Study, when
  - establishing spatial boundaries, and
  - considering regulatory and practical constraints for sampling.

CHAPTER 2
STEP 2. IDENTIFY THE DECISION



After reading this chapter you should know how to identify the principal study question and how to define options for addressing it (alternative actions).

2.1 Background

This step builds on the output of the previous step where you have:

- identified members of a planning team, including decision makers;
- concisely described the problem; and
- developed a conceptual model of the environmental problem to be investigated.

In Step 2 of the DQO Process, you should identify the key question that the study attempts to address and alternative actions that may be taken, depending on the answer to the key study question. Then you are able to combine these two elements to develop a decision statement. The decision statement is critical for defining decision performance criteria later in Step 6 of the Process.

In cases of multiple or complex problems, you should identify multiple decisions, organize the decisions sequentially (or logically), and examine the decisions to ensure consistency with the statement of the problem in Step 1. If the principal study question is not obvious and specific alternative actions cannot be identified, then the study may fall in the category of exploratory research, in which case this particular step of the DQO Process may not be needed.

2.2 Activities

In this step you should:

- identify the principal study question;
- define alternative actions;
- combine the principal study question and alternative actions into a decision statement and state each decision in terms of whether to take action. In some cases, this decision statement will be based on regulatory guidelines; and
- organize multiple decisions into an order of priority.

How do you identify the principal study question? Based on a review of the problem described in Step 1, you should identify the principal study question and state it as specifically as possible. A specific statement of the principal study question focuses the search for information needed to address the problem. The principal study question identifies key unknown conditions or unresolved issues that reveal the solution to the problem being investigated. EPA recommends that initially you should concentrate on only one principal study question and expand to other issues later. The following are examples of typical principal study questions:

- Does the concentration of contaminants in ground water exceed acceptable levels?
- Does the pollutant concentration exceed the National Ambient Air Quality Standard?
- Does a contaminant pose a human health or ecological risk?
- Is the contaminant concentration significantly above background levels (suggesting a release)?

In each case, the answer to the principal study question will provide the basis for determining the course of action that should be taken to solve the problem.

What are alternative actions and how should you define them? During this step, the planning team should identify the possible actions that may be taken to solve the problem, including an alternative that requires no action. The team should confirm that the alternative actions can resolve the problem (if it exists) and determine whether the actions satisfy regulations. An example of a principal study question and alternative actions is given in Table 2-1.

Table 2-1. An Example of a Principal Study Question and Alternative Actions

Principal Study Question	Alternative Actions
Are there significant levels of lead in floor dust at a children's residence.	Remove the children from the residence.
	Initiate a clean-up removal of lead-based paint.
	Take no action.

How do you develop a decision statement? After examining the alternative actions, you should combine the principal study question and alternative actions into a decision statement that expresses a choice among alternative actions. The following template may be helpful in drafting decision statements:

> Determine whether or not [unknown environmental conditions/issues/criteria from the principal study question] require (or support) [taking alternative actions].

Does the DQO Process address multiple decisions? If several separate decision statements must be defined to address the problem, you should examine how the decisions relate to one another and prioritize them in the order of the importance and sequence for resolving them. It may be helpful to document the prioritizing process proposed to resolve the problem using a diagram or a flow chart. An example is presented in Figure 2-1.

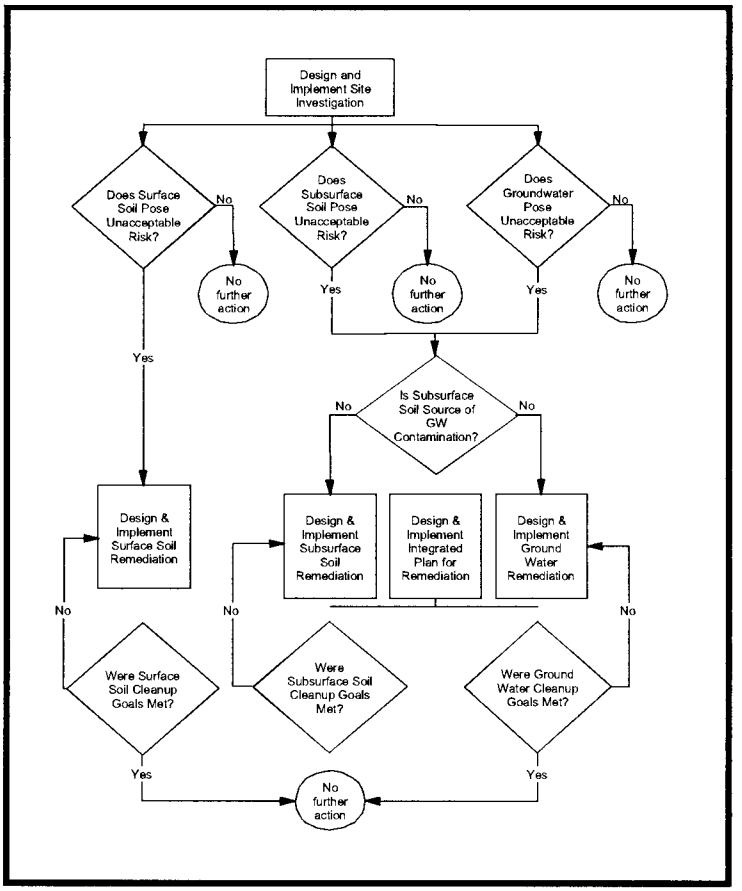


Figure 2-1. An Example of the DQO Process Applied to Multiple Decisions for a Hazardous Waste Investigation

2.3 Outputs

The output for this step is a decision statement that links the principal study question to possible actions that will solve the problem.

2.4 Examples

Example 1. Making Decisions About Incinerator Fly Ash for RCRA Waste Disposal

What was the Decision Statement? *The decision statement was determining whether waste fly ash was hazardous under RCRA regulations.*

What were the alternative actions? *If the waste was hazardous, disposal in a RCRA landfill was required. If it was not, the team decided that disposal in a sanitary landfill was acceptable.*

Example 2. Making Decisions About Household Dust for Lead Hazard Assessment

What was the Decision Statement? *The decision statement was determining if there were significant levels of lead in floor dust at the residence.*

What were the alternative actions? *If yes, the team planned follow-up testing to determine whether immediately dangerous contamination existed and the location of the contamination in the property. If no, the team decided there was not a potential lead hazard, and testing was discontinued.*

Example 3. Making Decisions About Urban Air Quality Compliance

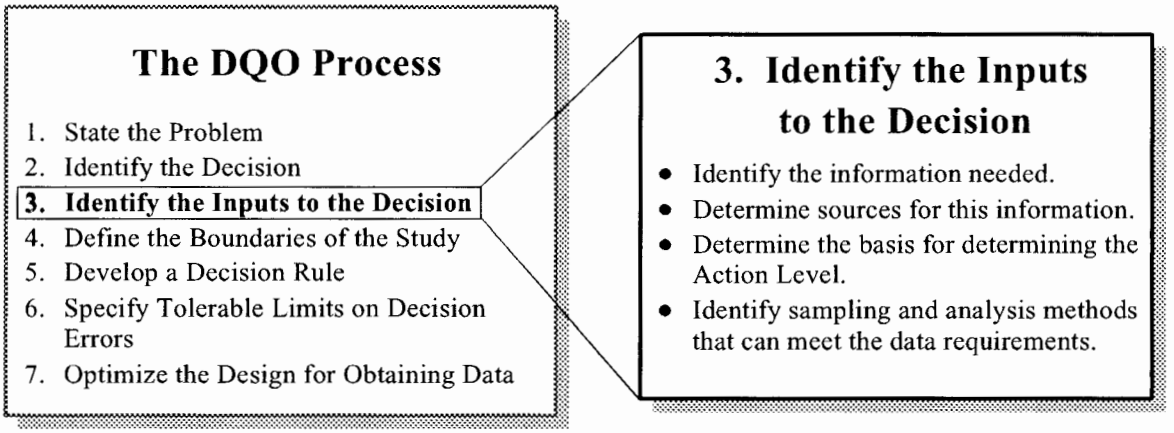
What was the decision statement? *The decision statement was determining if the PMSA of concern was in attainment for $PM_{2.5}$.*

What were the alternative actions? *If yes, monitoring was continued. If no, monitoring was continued and the $PM_{2.5}$ control strategies outlined in the State Implementation Plan (SIP) were implemented.*

Looking Ahead to other DQO steps:

- The principal study question will be used in constructing the baseline and alternative conditions in Step 6.
- Alternative actions will form the basis for determining the potential consequences of committing a decision error as addressed in Step 6.

CHAPTER 3
STEP 3. IDENTIFY THE INPUTS TO THE DECISION



After reading this chapter you should know the kinds of information that are required to investigate the problem and whether appropriate sampling and analytical methods are available.

3.1 Background

This step builds on the previous steps where you have:

- identified members of a planning team, including decision makers;
- concisely described the problem;
- developed a conceptual model of the environmental problem to be investigated; and
- identified the decision that needs to be made.

In Step 3 of the DQO Process you should identify the kind of information that is needed to resolve the decision statement and potential sources of this information (new data or existing data). This information should include the decision values (e.g., concentration of contaminants) information about its derivation. You should also determine if the appropriate analytical methodology exists to measure the environmental characteristics. Once you have determined what needs to be measured, you may refine the specifications and criteria for these measurements in later steps of the DQO Process.

3.2 Activities

In this step you should:

- identify the kinds of information needed;
- identify the sources of information;
- determine the basis for setting the Action Level; and
- confirm the appropriateness of proposed sampling and analyses methods.

How do you identify the kinds of information that you will need? You may identify information needs by asking the following questions:

- Is information on the physical properties of the media required?
- Is information on the chemical or radiological characteristics of the matrix needed?
- Can existing data be used to make the decision?
- Do we need to make new measurements of environmental characteristics?

If you decide that new measurements are needed, you should develop a list of characteristics that need to be measured to make the decision. For example, if the information can be obtained as an output from an environmental model (e.g., ground water transport), then the list of characteristics should include the inputs required for the model.

If the decision can be based on existing data, then the sources of these data should be examined to the extent possible to ensure that they are acceptable. If you consider integrating new data with existing data, parameters in the existing database need to be examined so that new samples can be collected (or analyzed) in a similar way and that the databases for new and existing data include common parameters. In some cases, statistical expertise is required to evaluate databases for possible aggregation because data collected for different purposes may not be compatible. For example, studies that model exposure to environmental contaminants may link environmental, toxicological, biological, geological, and census data. In these cases, issues such as physical properties of contaminants, environmental media, ingestion and inhalation rates, cancer slope factors, plant uptake rates, meteorological conditions, latitude, longitude, location of population centers and water bodies, and population density are inputs for evaluating exposure to the contaminant. Meta-data analysis offers the planning team options for using existing databases in conjunction with newly collected data. Existing data will also be evaluated quantitatively in Step 7, Optimize the Design for Obtaining Data.

How should you identify the source of the information? You should identify and document the sources for the information needed to resolve the decision. These sources may include results of previous data collections, historical records, regulatory guidance, professional judgment, scientific literature, or new data collections.

How do you determine the basis for the Action Level? The value for action is the threshold value (chosen in Step 5 of the DQO Process) that provides the criterion for choosing among alternative actions (e.g., whether to take action or not to take action or whether to choose action 1 versus action 2). Action Levels are concentrations of contaminants that are either based on regulatory requirements, based on risk assessments, based on performance criteria for analytical methodology (limitations of technology), or based on a reference standard. In this step, it is important for you to understand how the Action Level will be derived. In other words, you need to understand what information will be used to determine the Action Level, such as a promulgated regulation or a project-specific risk assessment. The actual numerical value of the Action Level need not be specified until DQO Process Step 5, Develop a Decision Rule, but a potential Action Level should be established. If the Action Level is based on a regulatory requirement, then the planning team will know the numerical value of the Action level at this step. However, if the Action Level is based on a risk assessment or other performance criterion, it may be best to defer the specification of the numerical value until after the study boundaries have been specified in DQO Process Step 4.

If the decision will be made relative to background concentrations (rather than a quantitative limit), then you should determine what constitutes background. Characteristics of the background need to be consistent with the characteristics of the area to be investigated. The actual numerical value of the Action Level will be established in Step 5, Develop a Decision Rule.

How should you identify that sampling and analysis methods that can meet the data requirements? Using the list of environmental characteristics that pertain to the decision, you should develop a list of sampling and analytical methods that may be appropriate for the problem being investigated. For example, you should specify sampling considerations (e.g., quantities) required for detecting analytes at low concentrations and procedures required to collect these sample quantities. You should also identify analytical methods that have appropriate detection limits (the minimum concentration that can be measured and reported with a specific confidence that the analyte concentration is greater than zero). Detection limits are analyte-, matrix- and instrument-specific. For example, atomic absorption spectroscopy or inductively coupled plasma emission spectrometry may not be sensitive enough to measure lead levels in water samples; however, graphite furnace atomic absorption spectroscopy would be capable of making these measurements.

Great importance should be given to the problem of minimizing bias as this is an important performance characteristic of sampling and analysis. The decision error rates to be established in Step 6 of the DQO Process rely on bias being kept to a minimum. Six major causes of bias have been identified for environmental sampling and analysis (1) non-representative sampling; (2) instability of samples between sampling and analysis; (3) interferences and matrix effects in analysis; (4) inability to determine the relevant forms of the parameter being measured; (5) calibration; and (6) failure to blank-correct. Some of the EPA methods are particularly subject to bias in calibration. For example, EPA methods for analyses of phenols in water exhibit around 50% bias due to calibration. Methods known to exhibit large biases should be avoided if possible.

Additional considerations include requirements for certification of personnel, and laboratory accreditation or Performance-Based Measurement Systems (PBMS). Laboratories analyzing environmental samples should follow standard protocols and procedures or use performance-based methods. When measurement requires the analysis of chemical, biological, or radioactive samples, it is advisable to select a laboratory that is accredited to perform the analyses. Requirements for accreditation include having qualified personnel, appropriate instrumentation, standard operating procedures, and proficiency in the analysis of samples for specific analytes or programs. For example, laboratories analyzing lead in paint, dust, and soil samples must be accredited through the National Lead Laboratory Accreditation Program (NLLAP) to become "EPA recognized." According to the Department of Housing and Urban Development's Guidelines (HUD, 1995), "property owners, risk assessors, inspector technicians, and contractors should ensure that laboratory analyses are performed by an 'EPA-recognized' laboratory;" a requirement also of EPA and many States.

3.3 Outputs

The outputs from Step 3 are:

- a list of environmental characteristics that will be measured to enable the planning team to make the decision;
- a list of information sources or methods that indicate how each Action Level will be derived;
- a list of information that may be applicable to uses of the data in future investigations [e.g., inputs to models, associated meta-data analysis (e.g., using latitude, longitude, census data) that may be appropriate to use for combining existing databases with newly collected data]; and
- confirmation that sampling and analytical methods exist (or can be developed) to meet the detection limit criteria required for collecting data, given the appropriate magnitude of the Action Level.

3.4 Examples

It is in this step that numerical quantities start making their appearance in the DQO Process.

Example 1. Making Decisions About Incinerator Fly Ash for RCRA Waste Disposal

Identify the kind of information. *To resolve the decision statement, the planning team decided to measure the cadmium concentration in the leachate resulting from Toxicity Characteristic Leaching Procedure (TCLP) extraction. Existing pilot study data provided information about variability, but there was not enough information to resolve the decision statement.*

Identify the source of information. *The Action Level was based on RCRA toxicity regulations for cadmium in TCLP leachate which is specified as 1.0 mg L.*

What sampling and analytical methods were appropriate? *Cadmium was measured in the leachate according to the method specified in 40 CFR 261, App. II. The detection limit was well below the Action Level.*

Example 2. Making Decisions About Urban Air Quality Compliance

Identify the kind of information. *To resolve the decision statement, the planning team obtained three years of $PM_{2.5}$ concentration measurements from the existing monitoring network within the PMSA of concern.*

Identify the source of information. *The 24-hr $PM_{2.5}$ federal standard of 65 $\mu g/m^3$ is attained when 98 percent of the daily concentrations, measured over three years, are equal to or less than the standard.*

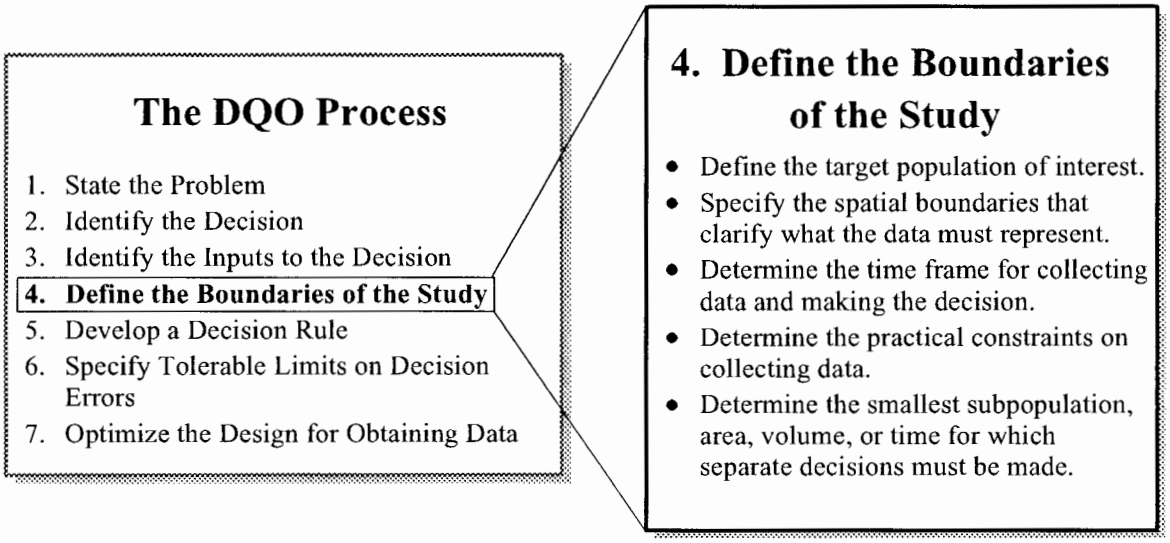
What sampling and analytical methods were appropriate? *The existing network consisted of three IMPROVE samplers, each equipped with a polytetrafluoroethylene (PTFE) membrane filter to collect aerosols for mass measurement. Gravimetry (electro-microbalance) was used as the method of quantitative analysis. The detection limit was well below the standard used for the Action Level.*

Looking Ahead to other DQO Steps:

- The effect of sampling methods (e.g., compositing) may affect the required detection limit and should be considered relative to analytical measurement methods. These issues are also considered in Steps 5 and Step 7.
- Criteria for existing data will be examined in Step 7, Optimize the Design for Collecting Data.
- Method detection limit and method quantitation limits identified in this step will be revisited in Step 7, Optimizing the Design for Collecting Data.

CHAPTER 4

STEP 4. DEFINE THE BOUNDARIES OF THE STUDY



After reading this chapter you should understand how to define the geographic and temporal boundaries of the problem, how to examine any practical constraints to collecting data, and factors that affect your selection of the unit for decision making.

4.1 Background

This step builds on the previous steps where you have:

- identified members of the planning team, including decision makers;
- concisely described the problem;
- developed a conceptual model of the environmental problem to be investigated;
- identified the decision that needs to be made; and
- identified sources of information, potential Action Levels, and possible measurement methods that are appropriate.

In Step 4 of the DQO Process, you should identify the target population of interest and specify the spatial and temporal features of that population that are pertinent for decision making.

It is difficult to interpret data that have not been drawn from a well-defined target population. The term "target population" refers to the total collection or universe of objects, or sampling units, to be studied and from which samples will be drawn. (In this context, the term

"sample" means the individual member, or unit, of the target population that is selected and measured or observed, such as a 100-gram scoop of soil, a cubic meter of air, a single fish, or single radiation measurement.) The term "sampling unit" is used in the more general and theoretical context when defining how the target population will be broken down into elementary components or members that can be selected and measured or observed. When the target population is made up of "natural units," such as people, plants, or fish, then the definition of a sampling unit is straightforward. However, many environmental studies involve target populations made up of continuous media, such as air, water, or soil. In this context, the sampling unit must be defined as some volume or mass to be selected which is often called the sample support (Myers, 1997). The actual determination of the optimal size of a sampling unit for environmental data collection efforts can be complicated, and usually will be addressed as a part of the sampling design in Step 7. Here in Step 4, the planning team should be able to provide a first approximation of the sampling unit definition when specifying the target population.

Quite often in environmental studies the target population is the set of all possible environmental samples (e.g., volume of soil, water, or air) that, taken together, constitute the geographic area of interest. The purpose of this step is to unambiguously define the spatial and temporal features of each environmental medium within a specific area or time period covered in the decision. A clear definition of the target population and its characteristics to the decision maker will make data interpretation more straightforward. The boundaries of the population include:

- spatial boundaries that define the physical area to be studied and generally where samples will be collected, and
- temporal boundaries that describe the time frame that the study will represent and when the samples should be taken.

You should use boundaries to ensure that the data collection design incorporates the time periods in which the study and decision should be implemented, areas where samples will be collected, and the time period to which the decision should apply. This should help you collect data that are representative of the population being studied. Defining boundaries before the data are collected can also prevent inappropriate combining of data sets in a way that masks useful information. The conceptual model that you developed in Step 1 of the DQO Process should provide essential input into defining the spatial boundaries.

Practical constraints that could interfere with sampling should also be identified in this step. A practical constraint is any hindrance or obstacle (such as fences, property access, water bodies) that may interfere with collecting a complete data set. These constraints may limit the spatial and/or temporal boundaries or regions that will be included in the study population and hence, the inferences (conclusions) that can be made with the study data.

As the final decision depends on data that are aggregated, you should carefully identify the size of "decision" units within which the data will be combined to make the decision. Factors

such as areas of potential risk, limits of remediation technology, future land uses, and activity patterns, may impact the size of the decision unit selected.

4.2 Activities

In this step you should:

- define the target population,
- determine the spatial and temporal boundaries,
- identify practical constraints, and
- define the scale of decision making.

How do you define the target population? It is important for you to clearly define the target population to be sampled. The target population is usually the set of all environmental samples about which the decision maker wants to draw conclusions. In a number of cases, defining the target population for an environmental study requires specifying the medium, such as ground water, ambient air, surface soil, etc. It may be helpful to "work backwards" and think of how you would define an individual sampling unit when trying to develop a clear definition of the target population.

How do you determine the spatial boundaries of the decision statement?

1. Define the geographic area applicable for the decision making.

You should define the entire geographic area where data are to be collected using distinctive physical features such as volume, length, width, or boundaries. Some examples of geographic areas are the metropolitan city limits, the soil within the property boundaries down to a depth of 6 inches, a specific water body, length along a shoreline, or the natural habitat range of a particular animal species. It is important to state as definitively as possible the media and geographic area; this statement may include soil depth, water depth, or distance inside a fence line. You should be careful when designating areas that are on the periphery of the geographic area because peripheral samples are subject to edge effects and contamination that is not associated with the spatial boundaries designated for the decision making. In Figure 4-1 the geographic area of the study has been indicated on a map in the area with a grid.

2. Divide the population into subsets that have relatively homogeneous characteristics.

You may consider dividing the target population into subpopulations that are relatively homogeneous within each area or subunit. When combined with an appropriate sampling design in Step 7, Optimize the Design for Obtaining Data, this approach can reduce the

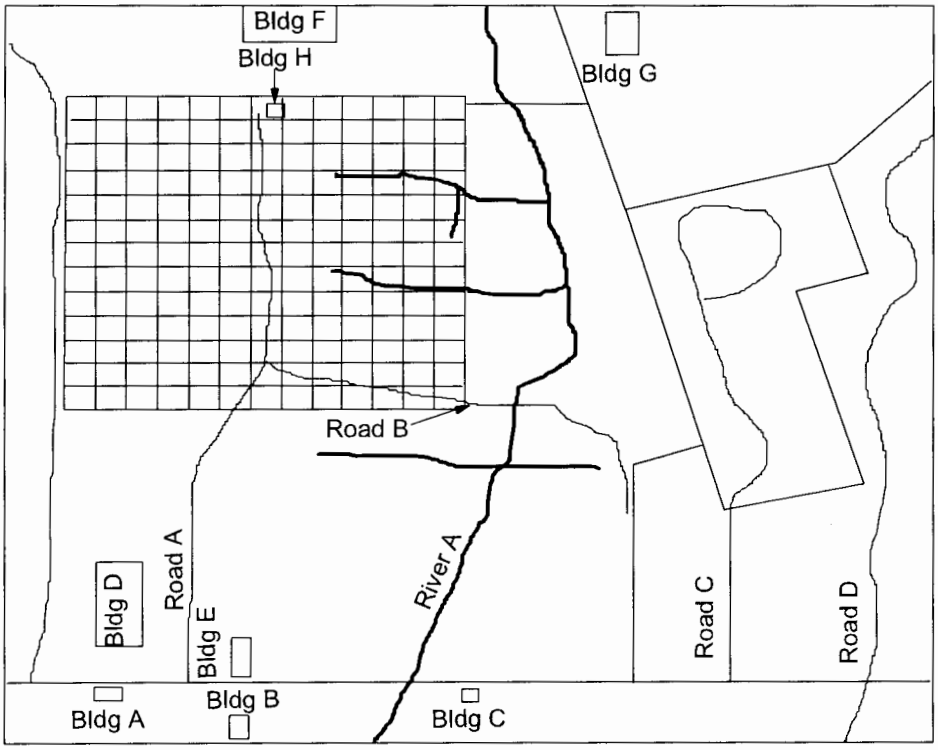


Figure 4-1. Geographic Boundaries Delineated Using a Map

number of samples required to meet the tolerable limits on decision errors (Step 6), and, thus, allow more efficient use of resources. It is often helpful to consider subdividing the target population in this way at this step because the planning team is focused on their understanding of how the target population's features and characteristics relate to the decision. The planning team can use its knowledge of the conceptual model (developed in Step 1, State the Problem) to consider how the characteristics of interest for the target population vary or change over space and time. This information will be useful when completing the subsequent activities in this step, and when considering alternative sampling designs (such as stratified random sampling) in Step 7, Optimize the Design for Collecting Data.

How do you determine the temporal boundaries of the decision statement?

1. Determine when to collect data.

Conditions may vary over the course of a study because of time-related phenomena such as weather conditions, seasons, operation of equipment under different environmental conditions, or activity patterns. Examples of these variations include seasonal ground

water levels, daily or hourly airborne contaminant levels in metropolitan areas, and fluctuations in pollutant discharges from industrial sources. These variations may impact the success of your data collection and the interpretation of data results. You should determine when conditions are most favorable for collecting data and select the most appropriate time period to collect data. For example, you may consider the measurement stability of the following:

- measurement of lead in dust on window sills may show higher concentrations during the summer when windows are raised and paint/dust accumulates on the window sill;
- terrestrial background radiation levels may change due to shielding effects related to soil dampness;
- measurement of pesticides on surfaces may show greater variations in the summer because of higher temperatures and volatilization;
- instruments may not give accurate measurements when temperatures are colder; or
- measurements of airborne particulate matter may not be accurate if the sampling is conducted in the wetter winter months rather than the drier summer months.

2. Determine the time frame for decision making.

It may not be possible to collect data over the full time period to which the decision will apply. This is particularly true for decisions that project future uses, such as "Brownfields" (an inactive property being put back into productive economic use after the relevant environmental agencies agree that contaminants once present at the property no longer pose an unacceptable risk to human health or to the environment). You should evaluate the population and determine the optimum time frame for collecting data, given that the medium may change over time, or the time constraints of the study relative to the decision making. You should specify if you are making a decision on whether the current medium meets a criterion, or if the medium will meet the criterion for some future time periods. You should define time frames for the overall population and for any subpopulation of interest; then address discrepancies that may arise from the short time frame of data collection relative to the long time periods for implementing decisions. For example, you may develop a statement for the decision to be based on:

- the condition of contaminant leaching into ground water over a period of a hundred years, or
- the risk conditions of an average resident over their average length of residence, which is estimated to be 8 years.

What kinds of practical constraints on collecting data should you identify? You should discuss the proposed data collection activities in light of any practical constraints that are related to the spatial or temporal boundaries of the study, to the availability of personnel, or to time and budgetary constraints (identified in Step 1 of the DQO Process). These constraints could include access to the property, availability and operation of equipment, and environmental conditions when sampling is not possible (high humidity, freezing temperatures). For example:

- it may not be possible to take surface soil samples beyond the east boundaries of a property under investigation because permission has not been granted by the owner of the adjacent property, or

- it may not be possible to collect dust wipe samples (for lead) if certified risk assessors are not available to supervise the sampling.

How do you define the scale of decision making? The scale of decision making refers to the way the planning team has delineated decision units and identified the smallest unit of area, volume, or time where data will be collected, analyzed, aggregated, and interpreted to make a decision and control decision error. The consequences of making incorrect decisions (Step 6) are associated with the size, location, and shape of the decision unit. It is important to consider present and future uses for the decision unit, where the decision unit is located (remote area versus densely populated area) and requirements for potential remediation. The consequences of a wrong decision (even if quite small) should be carefully considered. For example, if a decision, based on the data collected, results in a large land area being cleaned (soil removed to a certified disposal area) when the true conditions would not warrant a cleanup action, then the decision maker may have to incur a large cost unnecessarily. The area of land being sampled (decision unit) should be appropriate to the potential risk of an incorrect decision. When establishing the scale of decision making, take care that this is not so large that an incorrect decision could result in either an unacceptable resource expense or unacceptable threat to human health or the environment.

The question of using one large decision unit versus a number of small decision units is also an important consideration for the planning team. If there are many decision units and multiple decisions are made, then the team needs to consider whether they want to limit the probability of leaving *at least* one contaminated unit unremediated (rather than just *any* one unit). The chance of at least one incorrect decision increases exponentially. This is known as "comparison-wise" versus "experiment-wise" error rates. If multiple decisions are expected, and the planning team determines that the overall probability of making at least one decision error must be controlled, then consultation with a statistician is advisable.

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

- Risk – The scale of decision making based on risk is determined by the potential exposure an area presents; an individual unit of risk is called an exposure unit. For

example, in a study where the *decision statement* is, "Determine whether or not the concentration of lead in soil poses an unacceptable health risk to children and requires remediation," the *geographic area* is the top 6 inches of soil within the property boundaries, and the *population* is the collection of individual volumes of soil that could be selected for inclusion in a sample. The *scale of decision making* could be the size that corresponds to the area where children derive the majority of their exposure (such as a play area or an average residential lot size if the future land use will be residential). Studying the area at this scale will be protective of children, a sensitive population in risk assessment.

- Technological Considerations – A technological scale for decision making is defined as the most efficient area or volume that can be remediated with a selected technology. An example of a remediation unit would be the area of soil that can be removed by available technology under estimated working conditions if the decision will be made on the basis of bulldozer-pass-volume.

- Temporal Considerations – A temporal scale of decision making is based on exposure from constituents in media that change over time. For example, in order to regulate water quality, it would be useful to set a scale of decision making that reduces the time between sampling events. Using this scale the planning team could minimize the potential adverse effects in case the water quality changed between sampling events.

- Financial Scale – The financial scale is based on the actual cost to remediate a specified decision unit. For example, if a large exposure unit is identified, the costs of remediation could be prohibitive. In this case, the planning team may want to develop a different scale to narrow the data collection process and identify the distinct areas of contamination.

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

Further information on sampling designs and associated definitions on methods may be obtained from Gilbert (1987) and Thompson (1992).

4.3 Outputs

The outputs of this step are:

- detailed descriptions of the characteristics that define the population to be sampled,
- detailed descriptions of geographic limits (spatial boundaries) that are appropriate for the data collection and decision making,
- time frame appropriate for collecting data and making the decision,
- list of practical constraints that may interfere with the data collection, and
- appropriate scale for decision making.

4.4 Examples

Example 1. Making Decisions About Incinerator Fly Ash for RCRA Waste Disposal

What population was sampled? *Individual samples of fly ash that comprise the container were sampled and analyzed. The fly ash was not mixed with any other constituents except water (used for dust control). Each container of ash filled at least 70% of the waste container. In cases where the container was less than 70% full, the container was kept on-site until more ash was produced and the container was filled to capacity.*

What were the spatial boundaries? *Decisions applied to each container load of fly ash waste as the actual container made a natural physical boundary.*

What was an appropriate time frame for sampling? *The decision was to be based on the current concentration of cadmium in the waste fly ash. Contained in the containers, the waste did not pose a threat to humans or the environment. Additionally, since the fly ash was not subject to change, disintegration, or alteration, the decision about the waste characteristics was not influenced by temporal constraints. To expedite decision making, however, the planning team placed deadlines on sampling and reporting. The waste fly ash was tested within 48 hours of being loaded onto waste containers. The analytical results from each sampling round were completed and reported within 5 working days of sampling. The container was not used again until analysis had been completed and evaluated.*

What were the practical constraints for collecting data? *The most important practical constraint was the ability to take samples from the waste fly ash stored in the containers. Although the containers had open access, special procedures and methods based on EPA protocols were implemented so that samples were representative of the entire depth of the waste fly ash.*
What was the scale for decision making? *The decision unit was each container of waste fly ash.*

Example 2. Making Decisions about Urban Air Quality Compliance

What population was sampled? *The volume from samplers that represented fine particulate matter from urban air was sampled and analyzed.*

What were the spatial boundaries? *The spatial boundary was defined by the region represented by the PMSA of concern.*

What was an appropriate time frame for sampling? *The temporal boundaries had two components. Individual observations (i.e., daily concentrations) were based on 24-hour averages obtained each day of monitoring. The standard required that a decision be made, and subsequent action taken, after 3 years of data collection. Monitoring results were assumed to characterize both the near past (i.e., previous 3 years) and current air quality, unless substantial upward or downward trends were observed in daily* $PM_{2.5}$ *concentrations.*

What were the practical constraints for collecting data? *Given that the monitoring network and sampling plan were already established, the only potential practical constraint was the continual operation of the monitoring network. If a monitor became defective, the planning team decided to either collect a smaller sample size (number of samples) over the 3-year period, or to extend the period for collecting data to obtain the required number of samples.*

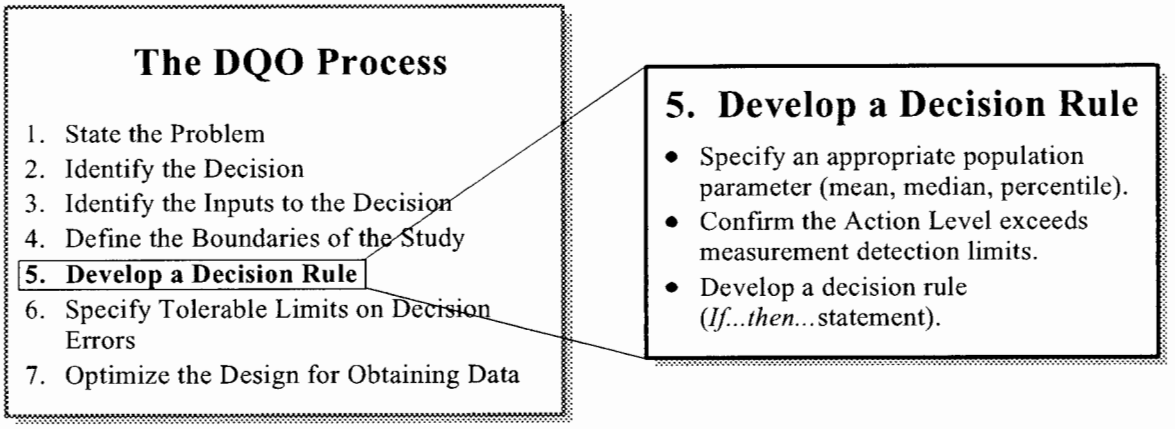
What was the scale for decision making? *The decision unit was the geographic region represented by the PMSA over the 3-year period of data collection.*

Looking ahead to other DQO steps:

- The way in which you divide the problem into strata may affect the number of samples required to meet the tolerable limits for decision errors specified in Step 6.
- The scale of decision making may have an impact on the performance criteria and the consequences of decision errors in Step 6.
- Outputs from Step 4 may potentially affect the sampling design developed in Step 7.

CHAPTER 5

STEP 5. DEVELOP A DECISION RULE



After reading this chapter you should know how to construct a theoretical "If...then..." decision rule that defines how the decision maker would choose among alternative actions if the true state of nature could be known with certainty.

5.1 Background

This step builds on the previous steps where you have:

- identified members of the planning team, including decision makers;
- concisely described the problem;
- developed a conceptual model of the environmental problem to be investigated;
- identified the decision that needs to be made;
- identify sources of information, potential Action Levels, and possible measurement methods that are appropriate; and
- decided on the spatial/temporal boundaries of the decision.

In Step 5 of the DQO Process, you should imagine that perfect information will be available for making decisions. Under the assumption that there is no uncertainty in the decision making process, the planning team integrates the outputs from previous steps with inputs developed in Step 5 into an unambiguous "*If...then...*" statement (*theoretical* decision rule). This rule describes the conditions under which possible alternative actions would be chosen.

You need to conduct the following activities in this step:

- specify the population parameter (e.g., mean, median, percentile, or total amount) that the DQO planning team considers to be important to make decisions about the target population;

- choose an Action Level (if not already established) that sets the boundary between one outcome of the decision process and another outcome;

- select the measurement and analysis methods capable of performing over the expected rate of values and verify that the Action Level is greater than the detection limit of the measurement method that will be used; and

- construct the theoretical "*If...then...*" decision rule by combining the *true* value of the selected population parameter and the Action Level (from above) with the scale of decision making (from Step 4) and the alternative actions (from Step 2). This decision rule will state the alternative actions that would be taken depending on the true value of the parameter relative to the Action Level.

Note that the "*If...then...*" decision rule is a *theoretical* rule because it is stated in terms of the *true* value of the population parameter, even though in reality the true value is never known. In practice, the decision is made by using an *operational* decision rule that uses an *estimate* (based on the actual data) of the true value of the population parameter. The reason for specifying the *theoretical* rule is to focus the attention of the DQO planning team on how they would make decisions if they had perfect knowledge of the population. This helps clarify what the team really wants to know to support the decision. In Step 7 of the DQO Process, the planning team will select the operational decision rule they believe will most efficiently meet the requirements specified in the first six steps of the DQO process.

5.2 Activities

In this step you should:

- define the population parameter;
- determine what action is needed; and
- confirm that the Action Level exceeds minimum detection limits.

What population parameter best characterizes the population of interest? In this step you should select a population parameter (such as the true mean, median, percentile, or total amount) that summarizes the critical characteristic or feature of the population that will be compared to the Action Level to make a decision. In some cases, the parameter that must be used may be specified in a regulation. In other cases, the DQO planning team will select the parameter based

on specific needs and considerations. A comparison of the different population parameters and their application to a decision rule is presented in Table 5-1.

Table 5-1. Population Parameters and Their Applicability to a Decision Rule

Parameter	Definition	Example of Use
Mean	Average	Central tendency: Comparison of middle part of population to Action Level. Appropriate for chemical that could cause cancer after a long-term chronic exposure. Use of the mean and the total amount of media (e.g., mass of soil or water) allows a planning team to estimate the total amount of a contaminant contained in the soil or water body. The mean is greatly influenced by extremes in the contaminant distribution, and not very useful if a large proportion of values are below the detection limit.
Median	Middle observation of distribution; 50th percentile; half of data is above and half is below	Better estimate of central tendency for a population that is highly skewed (nonsymmetrical). Also may be preferred if the population contains many values that are less than the measurement detection limit. The median is not a good choice if more than 50% of the population is less than the detection limit because a true median does not exist in this case. The median is not influenced by the extremes of the contaminant distribution.
Percentile	Specifies percent of sample that is below the given value; e.g., the 80th percentile should be chosen if you are interested in the value that is greater than 80% of the population.	For cases where only a small portion of the population can be allowed to exceed the Action Level. Sometimes selected if the decision rule is being developed for a chemical that can cause acute health effects. Also useful when a large part of the population contains values less than the detection limit. Often requires larger sample sizes than mean or median.

It must be noted, however, that the more complex the parameter chosen, the more complex will be the decision rule and accompanying data collection design. The most common parameter used in decision making is the population mean because the mean is frequently used to model random exposure to environmental contamination. Aside from scientific or policy considerations, the mathematical and statistical properties of the mean are well understood. You should consult a statistician if you are uncertain as to the choice of an appropriate parameter.

What Action Level is needed for the decision? In addition to specifying the population parameter, you will need to specify the Action Level that will be used to choose between courses of action. For example, the decision maker may take one action if the true value of the parameter exceeds a specified value (Action Level) and a different action otherwise. There are basically two kinds of Action Levels – those predetermined and those determined during the DQO Process.

Examples of predetermined Action Levels are fixed standards such as drinking water standards or technology-based standards. For example, in the area of childhood lead poisoning

prevention, EPA's Office of Pollution Prevention and Toxics has proposed hazard levels for lead in residential dust and soil to protect children from significant lead exposures (40 CFR 745). Also, in the area of air quality control, EPA's Office of Air and Radiation has promulgated National Ambient Air Quality Standards for priority pollutants such as carbon monoxide, lead, nitrogen dioxide, ozone, particulate matter (PM_{10}), and sulfur dioxide, as well as other pollutants that include fine particulate matter ($PM_{2.5}$) (40 CFR 50).

Examples of investigation-specific Action Levels are background standards or specific risk-based standards. For the case of investigation-specific Action Levels, one consideration in selecting the Action Level is its degree of conservatism, i.e., whether the level is a very low value or a higher value. You will need to decide whether to set the Action Level at a threshold of real concern, or at a lower (more conservative) value that, if exceeded to some degree, may not necessarily pose a serious risk. A more conservative Action Level may require a more sensitive analytical method that has appropriate detection limits.

Does the Action Level exceed measurement detection limits? You will need to determine the detection limit for each potential measurement method identified in Step 3. If the detection limit for a measurement method exceeds the Action Level, then a more sensitive method should be specified or a different approach should be used.

Detection limits are defined specific to an intended purpose. The DQO planning team should choose the definition that is most appropriate to the "*If...then...*" decision rule being used. For example, if the decision rule is used to decide if a contaminant exists at the study site, then the detection limit should be one that provides for a high probability of positive identification and presence in the matrix and a low probability of false confirmation. However, if the decision rule is used to compare a mean to a threshold action level, then the detection limit should be defined in terms of the reliability of quantitation.

5.3 Outputs

After you have completed the above activities, you can construct the theoretical "*If...then...*" decision rule by combining the selected population parameter and Action Level with the scale of decision making (from Step 4) and the alternative actions (from Step 2). An example of a theoretical decision rule is:

> *If* the true mean dioxin concentration in the surface 2 inches of soil of a decision unit (20 ft by 100 ft) exceeds 1 ppb, *then* remove a 6 inch layer of soil. *If* the true mean is not greater than 1 ppb, *then* do nothing.

5.4 Examples

Example 1. Making Decisions About Incinerator Fly Ash for RCRA Waste Disposal

What was the decision rule and Action Level? *The planning team was interested in the true mean concentration of cadmium in the TCLP leachate for each container. If the true mean concentration of cadmium from the fly ash leachate in each container load was greater than 1.0 mg/L, then the waste was considered hazardous and disposed of at a RCRA landfill. If the true mean concentration of cadmium from the waste fly ash leachate was less than 1.0 mg/L, then the waste was considered nonhazardous and disposed of in a sanitary landfill.*

Example 2. Making Decisions About Urban Air Quality Compliance

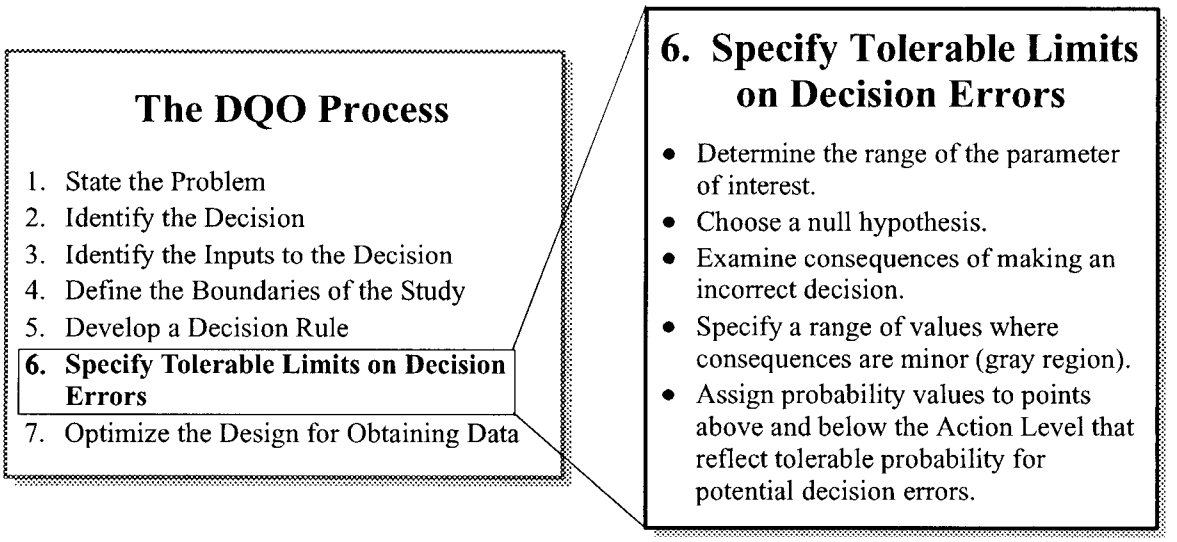
What was the decision rule and Action Level? *The population parameter of interest that characterizes $PM_{2.5}$ air quality was the true long-run proportion of daily concentrations falling below the 24-hr $PM_{2.5}$ federal standard of 65 $\mu g/m^3$. If the true proportion of daily concentrations less than or equal to 65 $\mu g/m^3$ was greater than or equal to 0.98, then the local region was considered in attainment for $PM_{2.5}$, so monitoring was continued, but no other action was taken. If the true proportion of daily concentrations less than or equal to 65 $\mu g/m^3$ was less than 0.98, then the local region was considered in nonattainment for $PM_{2.5}$, so monitoring was continued and the $PM_{2.5}$ control strategies outlined in the State Implementation Plan were implemented.*

Looking ahead to other DQO steps:

- Step 6 provides key information that will be used with the outputs of this step to select the sampling and analysis methods.
- Step 6 addresses the questions of what risk of an incorrect decision can be tolerated.

CHAPTER 6

STEP 6. SPECIFY TOLERABLE LIMITS ON DECISION ERRORS



After reading this chapter you should understand why specifying tolerable limits on decision errors is required to continue the DQO Process and the meaning of the concepts and terms used in completing this task. You should be able to specify tolerable limits on decision errors for your problem.

6.1 Background

This step builds on the previous steps where you have:

- identified members of the planning team, including decision makers;
- concisely described the problem;
- developed a conceptual model of the environmental problem to be investigated;
- identified the decision that needs to be made;
- determined the type of information required, the Action Level, and probable measurement methods;
- decided on the spatial/temporal boundaries of the decision; and
- decided on the theoretical *"if... then"* decision rule.

In Step 6 of the DQO Process you no longer imagine that perfect information on unlimited data will be available for making decisions as you did in Step 5. You now face the reality that you will not have perfect information upon which to base your decisions. Instead you will be making decisions based on a set of sample data subject to various errors which is only part of the much larger population of interest. Inherent in the use of sampled data for making decisions is the fact that those decisions can, and occasionally will, be wrong. In this step of the DQO Process, numerical values will be considered in an attempt to keep the possibility of a decision error to a minimum.

The purpose of Step 6 is to specify quantitative performance goals for choosing between the two alternative actions decision rule. These goals are expressed as probabilities of making errors in your decision at selected true values of the parameter of interest specified in Step 5. These decision performance goal probabilities are a statement of the amount of uncertainty you are willing to tolerate in your decisions at a few specific critical true values of the parameter of interest.

6.2 Activities

You should conduct the following activities in Step 6:

- determine the sources of error in the sample data set;
- establish a plausible range of values for the parameter of interest;
- define the two types of potential decision errors and the consequences of making those errors;
- determine how to manage potential decision errors;
- select the baseline condition of the environment that will be assumed to be true in the absence of overwhelming evidence to the contrary;
- specify a range of possible parameter values where the consequences of a false acceptance decision error are considered tolerable (gray region); and
- assign probability values at several true value points above and below the Action Level that reflect your tolerable probability for the occurrence of decision errors.

What are sources of error in the sample data set? A decision error occurs when the sample data set misleads you into making the wrong decision and, therefore, taking the wrong response action. The possibility of a decision error exists because your decision is based on sample data that are incomplete and never perfect. Even though the data collection method and analysis method may be unbiased, the sample data are subject to random and systematic errors at different stages of acquisition, from field collection to sample analysis. The combination of all these errors is called "total study error." There can be many contributors to total study error, but there are typically two main components:

- Sampling design error – This error is influenced by the inherent variability of the population over space and time, the sample collection design, and the number of

samples. It is usually impractical to measure the entire decision unit, and limited sampling may miss some features of the natural variation of the measurement of interest. Sampling design error occurs when the data collection design does not capture the complete variability within the decision unit to the extent appropriate for the decision of interest. Sampling design error can lead to random error (i.e., variability or imprecision) and systematic error (bias) in estimates of population parameters.

- 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, data reduction, transmission, and storage.

Total study error directly affects the probability of making decision errors. Therefore, it is essential for you to manage total study error by your choice of sample design and measurement system. This will enable you to control the possibility of making decision errors to acceptable levels. Figure 6-1 shows an example of how total study error (also known as Total Variability) can be broken down further into components that will relate to the data collection process.

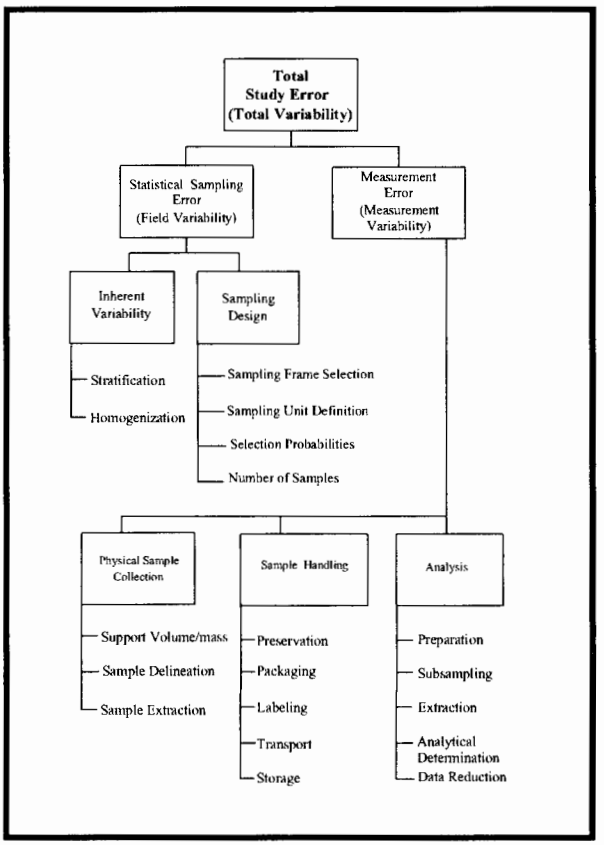


Figure 6-1. An Example of How Total Study Error Can Be Broken Down by Components

How do you establish a plausible range of values for the parameter of interest? You should establish a plausible range of values for the parameter of interest by approximating its upper and lower bounds based on currently available information, professional judgment, or historical data. This helps focus the process of defining probability limits on decision errors only on the relevant values of the parameter. For example, if the parameter of interest is a mean, the range might be defined using the lowest and highest concentrations at which the contaminant is thought to exist at the property. This range of values is useful when discussing the Decision Performance Goal Diagram (to be discussed later).

How are decision errors defined? If perfect knowledge of the true value of the parameter of interest in a decision unit were available to you, you could simply apply the theoretical decision rule from Step 5 to the known true value, make your decision, and not be concerned with decision errors. However, in real life you use sample data to make the decision and, consequently, the chance of a decision error becomes reality.

Due to the uncertainty inherent in decisions based on sample data, it is possible to get results that will not clearly tell you if the true value is below the Action Level or above the Action Level. It becomes necessary to label one of these two possibilities as the baseline condition so that a decision can still be made in these situations. The baseline condition then becomes the *de facto* decision outcome when there is insufficient evidence to refute it and the other condition then becomes the alternative decision. For example, in legal decisions on human behavior, the baseline condition is "innocent until proven guilty."

In environmental decisions affecting human health and the environment, the baseline condition is more flexible and may depend on your situation. In certain instances, the baseline condition for your problem may be prescribed for you in regulations. For example the baseline condition in RCRA facility monitoring is that the concentrations in ground water are less than or equal to the background concentrations. If the baseline condition is not specified for you, you must select it based on careful consideration of the consequences of making decision errors and taking the wrong actions. This selection may be based on your conceptual model for the decision unit, i.e., based on prior information, you have good cause to think that the true value for the decision unit is above the Action Level.

The probabilities of making decision errors with sample data can be quantified through the use of a statistical decision procedure known as hypothesis testing. When hypothesis testing is applied to decision making, the sample data are used to choose between a baseline condition of the environment and an alternative condition. The test can then be used to show either that there is insufficient evidence to indicate that the baseline condition is false (and therefore you accept the default that the baseline condition is presumed to be true), or that the baseline condition is probably false (and therefore the alternative condition is probably true). The burden of proof is placed on rejecting the baseline condition. This approach is taken 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. It is *critical* to understand that selection of the baseline condition is important to the outcome of the decision process. The exact same set of sample data from a decision unit can lead to different decisions depending on which possibility was chosen as the baseline condition.

A false rejection decision error[1] occurs when the limited amount of sample data lead you to decide that the baseline condition is probably false when it is really true. In the reverse case, a

---

[1]In previous editions of *Guidance for Data Quality Objectives Process (EPA QA/G-4)* (U.S. EPA, 1994a), false rejection was called "false positive" and false acceptance was called "false negative."

false acceptance decision occurs when the sample data lead you to decide that the baseline condition is probably true when it is really false. To understand these definitions you may find it helpful to note that an acceptance decision is to decide the baseline condition is true and a rejection decision is to decide the alternative condition is true. Hence, a false rejection decision incorrectly decides that the alternative is true, and a false acceptance decision incorrectly decides that the baseline is true (see Table 6-1). For example, suppose you strongly believe that the true value of the parameter of interest exceeds the Action Level (i.e., the baseline condition states that the true value of the parameter of interest exceeds the Action Level). If your baseline assumption is actually correct and the sample data, by chance, contain an abnormally large proportion of low values, you would conclude that the true value of the parameter of interest does not exceed the Action Level. In reality, the true value did exceed the Action Level; therefore, you would then be making a false rejection decision error.

Table 6-1. False Acceptance and False Rejection Decisions

Decision Based on Sample Data	**True Condition**	
	Baseline is True	Alternative is True
Decide baseline is true	Correct Decision	*Decision Error (False Acceptance)*
Decide alternative is true	*Decision Error (False Rejection)*	Correct Decision

Another example would be a regulatory situation in which an effluent discharge should not exceed the permitted level. Your baseline condition would be that the true parameter value of the effluent is less than or equal to the permitted level; your alternative would be that the true parameter exceeds the permitted level. If the baseline condition was actually correct, but your sample data happened to have a preponderance of high values, you could conclude the effluent exceeds the permitted level. This would be a false rejection decision error and is sometimes called a false positive decision error. The reverse (a false acceptance decision error) is sometimes called a false negative decision error.

In the statistical language of hypothesis testing, the baseline condition is called the null hypothesis (H_o) 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. Statisticians label a false rejection decision error as a Type I error and the measure of the size of this error (probability) is labeled alpha (□), the hypothesis test's level of significance. Statisticians label a false acceptance decision error as a Type II error; the measure of the size of this error (probability) is labeled beta (□). Both alpha and beta are expressed numerically as probabilities. The statistical power of a test of hypothesis is equal to 1-□.

How can you manage potential decision errors? Although the possibilities of making decision errors can never be eliminated totally, you can manage them. To manage the possibilities of

decision errors, your planning team focuses mostly on the largest components of total study error. If the sampling design error is believed to be relatively large, you can manage the chance of making a decision error by collecting a larger number of samples or developing a better sampling design, i.e., a better way of deciding where and when to sample. If the analytical component of the measurement error is believed to be relatively large, you can manage it by analyzing multiple individual samples, or by using more precise and accurate analytical methods. In some instances your planning team will actually be able to address both components of total error.

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 for you to make a defensible decision based on relatively imprecise data or on a small amount of data. For example, in the early phases of a hazardous site assessment, the consequences of deciding that an area of a site is hazardous, when in reality it is not, may be relatively minor. In this case, you 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 (i.e., human health effects), you will want to develop a data collection design that exercises more control over sampling design and measurement error. For example, in a waste discharge investigation, deciding that a discharge is not hazardous when it truly is hazardous may have serious consequences because the discharge may pose a risk to human health and to the environment. Therefore, the decision made during this phase of the investigation may need to be supported by a large amount of data and analyzed using very precise and accurate analytical methods.

You will need to balance the consequences of decision errors against the cost of limiting the possibility of these errors. 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 decisions with potential consequences should be explored fully by your 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 option, 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 your planning team during their activities under Step 7 of the DQO Process.

Decision errors can also occur that are independent of the use of statistical hypothesis testing. An example could be that the data were manipulated prior to use in decision making by an outside agent censoring the reported values. This is sometimes found in the collection of screening data where insufficient training on the importance of adherence to QA protocol and practice has resulted in data being recorded in an erroneous fashion. If data has been manipulated prior to use in decision making, the assumed false rejection and false acceptance error rates become invalid.

How can you represent the quality of a decision process? There is a graphical construct called a Decision Performance Curve that represents the quality of a decision process. In statistical hypothesis testing usage, an operating characteristic curve or a power curve serve similar purposes. Figure 6-2 depicts an example Decision Performance Curve and shows the range of possible true values of the parameter of interest (including the Action Level) decided in Step 6, on the x-axis and the range of probabilities (0 to 1) of deciding that the parameter of interest exceeds the Action Level along the y-axis. Intuitively, the probability of deciding the parameter of interest exceeds the Action Level is small for low true values and increases as the true value increases. A full Decision Performance Curve is actually a continuous curve from the lowest true value to the highest true value. If you had perfect knowledge of the true value of the parameter of interest, a Decision Performance Curve would have a probability of 0 for any true value less than the Action Level and jump to a probability of 1 for any true value above the Action Level. Since you are dealing with sampled data (containing error), the probabilities will more realistically increase gradually from near 0 for true values far below the Action Level, to near 1 for true values far above the Action Level. The shape and steepness of this curve is a consequence of the sample design and number of samples taken.

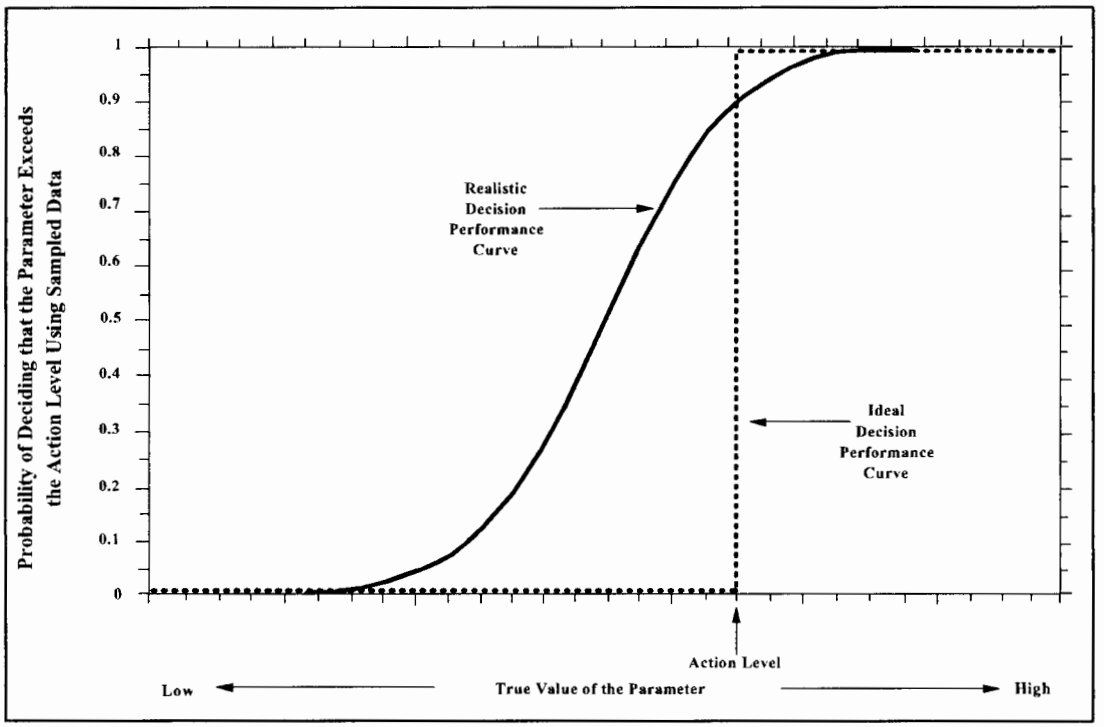


Figure 6-2. An Example of a Decision Performance Curve

The following subsections describe the process of selecting a baseline condition, defining a gray region, and establishing Decision Performance Goals (DPGs) by stating tolerable decision error probabilities at a few critical true values of the parameter of interest. The combined

information from these activities can then be displayed graphically as a Decision Performance Goal Diagram (DPGD) that approximates the Decision Performance Curve. This DPGD stipulates your tolerable risks of decision errors and allows you to communicate them to others, including your sample design team and all stakeholders. The Decision Performance Curve is then the overlay on the diagram and can be used to assess performance.

How do you select the baseline condition? If your baseline is not established by regulatory considerations, your planning team should define the baseline condition based on the relative consequences of the decision errors.

The baseline condition is the one that will be kept until overwhelming evidence (in the form of data to be collected) is presented to make you reject the baseline condition in favor of the alternative. You should use your 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, you would judge the threat to public health against spending unnecessary resources.

Define the baseline condition and the alternative condition and assign the terms "false rejection" and "false acceptance" to the appropriate decision error. An alternative name for "false rejection" is "false positive" or Type I Error (by statisticians principally). The alternative name for "false acceptance" is "false negative" or Type II Error. 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.

You should designate the areas above and below the Action Level as the range where the two types of decision errors may occur. This activity has two steps:

1. Define both types of decision errors and establish the "true state of nature" for each decision error. The "true state of nature" is the actual condition of the parameter of interest in the decision unit which is unknown to the decision maker. You should state both decision errors in terms of the parameter of interest, the Action Level, and the alternative actions.

2. Specify and evaluate the potential consequences of each decision error. For example, the 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 to the environment. Conversely, the 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 for further study.

   You 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 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.

How do you specify a range of possible true parameter values where the consequences of a false acceptance decision error are considered tolerable (gray region)? The gray region is one component of the quantitative decision performance criteria that is specifically used to limit impractical and nonfeasible number of samples. The gray region is a range of true parameter values within the alternative condition near the Action Level where it is "too close to call." This gray region is where sampled data may correctly reject the baseline condition, but the sampled data frequently do not provide sufficient evidence to be overwhelming. In essence, the gray region is an area where it is not considered feasible to control the false acceptance decision error limits to lower levels because the high costs of sampling and analysis outweigh the potential consequences of choosing the wrong course of action (see Figure 6-3 for example).

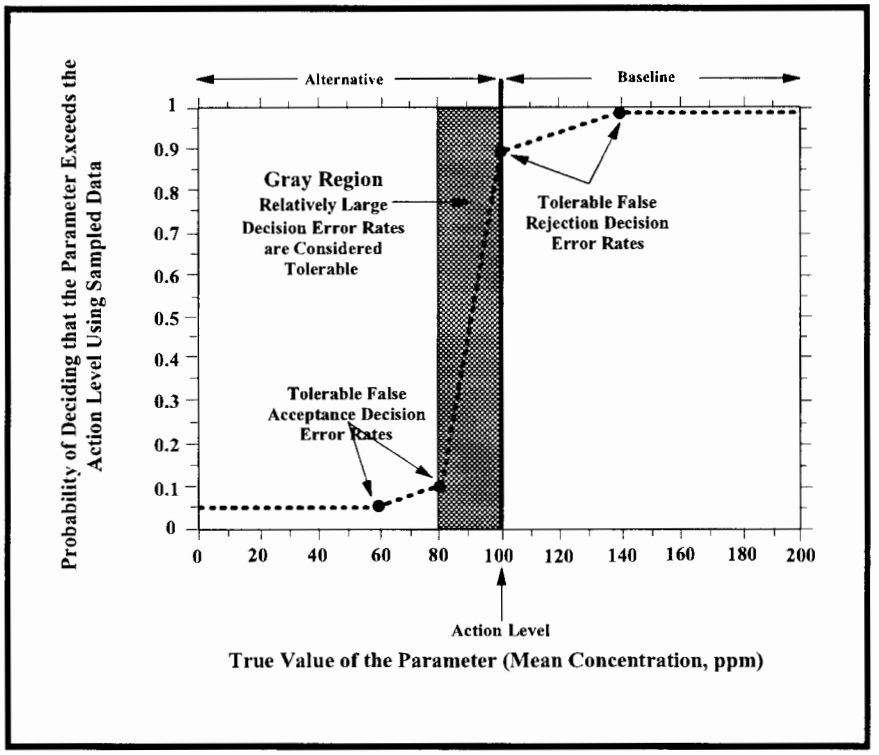


Figure 6-3. An Example of a Decision Performance Goal Diagram (Baseline Condition: Parameter Exceeds the Action Level)

The first boundary of the gray region is the Action Level itself. Your planning team establishes the other boundary of the gray region 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 low limit on the probability of this decision error occurring.

For example, suppose the baseline condition is that the true mean level of contaminant does not exceed 1.0 mg/L and the result of a sample of five observations reveals a sample mean of 1.05 mg/L. Is this sufficient evidence to reject the baseline condition? If the natural variability of the contaminant was low, then probably this would be enough evidence. If the natural variability was quite high (i.e., a coefficient of variation of 50%), then the evidence would not be overwhelming, as a result of this happening quite naturally. On the other hand, if the sample mean had been 1.50 mg/L, even high variability could not hide the fact that the baseline condition had been exceeded. The second boundary of the gray region is that value that you decide represents overwhelming evidence to reject the baseline condition.

In general, the narrower the gray region, the greater the number of samples needed to meet the criteria because the area of uncertainty has been reduced. The width of the gray region may be wide during early phases of the study process, but narrowed at later stages to determine if the parameter of interest is only slightly different than the Action Level.

In statistical hypothesis testing language, the width of 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 many calculations for determining the number of samples that need to be collected so that you will have your stated confidence in decisions made based on the data collected.

How do you 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 when making the decision using sampled data. 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 of making an incorrect decision.

At a minimum, you 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 (large probabilities). 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. 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, your 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.

The value of 0.01 should not be considered a prescriptive value for setting decision error rates, nor should it be considered EPA policy to encourage the use of any particular decision error rate. Some programs, for example Superfund, give guidance on alternative values for starting points. In the *Soil Screening Guidance: User's Guide* (U.S. EPA, 1996), the starting value for false rejection is 0.05, and for false acceptance, 0.20. The actual values finally selected by the planning team will depend on the specific characteristics of the problem being investigated.

Figures 6-3 and 6-4 illustrate some key outputs of Step 6 of the DQO Process for an example, but with opposite baseline conditions and different project specific-considerations. The DPGD is a special schematic representation of a Decision Performance Curve. While the Decision Performance Curve is a continuous curve, the schematic DPGD depicts only a few critical points on that curve. These few points represent your tolerable error limits, or DPGs, at a few critical values. Your sampling design team will use this as the criteria for any sampling plan they design. 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. You are answering the question at several selected true values of the parameter of interest:

> If the true value of the parameter of interest were at this level, how strong of an aversion would I have if the data misled me into making the wrong decision and taking action?

Figure 6-3 shows the case where a decision maker considers the more severe decision error to occur above the Action Level and has labeled that as baseline. Figure 6-4 shows the reverse, the case where the decision maker considers the more severe decision error to occur below the Action Level.

Consider Figure 6-3 where the baseline condition is that the parameter exceeds the Action Level (in statistical terms, H_0: the parameter equals or exceeds the Action Level and H_A: the parameter is less than the Action Level). The plausible range of values based on professional judgment was from the Detection Limit (as the Detection Limit was 0.01, it is essentially zero for purposes of the DPGD) to 200 ppm. The Action Level was 100 ppm (from the permit for this investigation). A false rejection would be saying the parameter is less than the Action Level, when, in fact, it is really greater. A false acceptance would be saying the parameter level is above the Action Level, when, in reality, it is below the Action Level. The gray region is the area where you consider it is tolerable to make a decision error as it is "too close to call." For example,

suppose you decided the true parameter level was above the Action Level (100 ppm) when in reality it was 99 ppm. Although an error has occurred (false acceptance), it is not particularly severe because the difference of 1 ppm on human health and financial resources is minimal. On the other hand, suppose you decided the true parameter level was above the Action Level (100 ppm) when in reality it was 80 ppm. Again, an error has occurred (false acceptance), but it is severe because a difference of 20 ppm is quite considerable. In this particular case the planning team chose 80 ppm as the edge of their gray region because it represented the case where errors in decision making have a great impact on resources. The planning team then assigned risk probabilities to the chance of making decision errors for various true values of the parameter. They agreed that, if the true value was 80 ppm and they decided (from the data yet to be collected) that the true value exceeded 100 ppm, they were only willing to accept a 10% risk of this happening. The team then considered the implications of what adverse effect would occur if the true value was 60 ppm, but they decided the parameter was greater than 100 ppm. The analysis showed a huge expenditure of resources, so the planning team elected to take only a 5% risk of this happening. They did a similar exercise with the tolerable false rejection error rates.

Now consider Figure 6-4 where the baseline condition is that the parameter is less than the Action Level (in statistical terms, H_0: the parameter is less than or equal to the Action Level and H_A: the parameter is greater than the Action Level). Notice how the DPGD looks very similar to that of Figure 6-3, except that the gray region is on the other side of the Action Level, and false rejection and false acceptance have now been switched. In statistical terms, this is because a false rejection is defined as rejecting H_0 when H_0 is really true, and false acceptance to be accepting H_0 when H_0 is really false.

Figure 6-4 shows that at the Action Level the decision maker will tolerate a 10% chance of deciding that the true value is below the Action Level when it is really above the Action Level. If the true value is 140 ppm, the decision maker will tolerate only a 1% chance of deciding the true value is below the Action Level when it is really above the Action Level. At the edge of the gray region, 120 ppm, the decision maker is willing to tolerate a 10% risk of saying it is above the Action Level when it is really below the Action Level. At 60 ppm, the decision maker is only willing to tolerate a 5% risk of a decision error. These probabilities represent the risk to the decision maker of making an incorrect decision for the selected true values.

6.3 Outputs

The outputs from this step are your baseline condition, your gray region, and your set of tolerable decision error limits at selected true values of the parameter of interest. These selections are based on a consideration of the consequences of making incorrect decisions. The baseline condition, the gray region, and your tolerable limits on decision errors are summarized in a Decision Performance Goal Diagram.

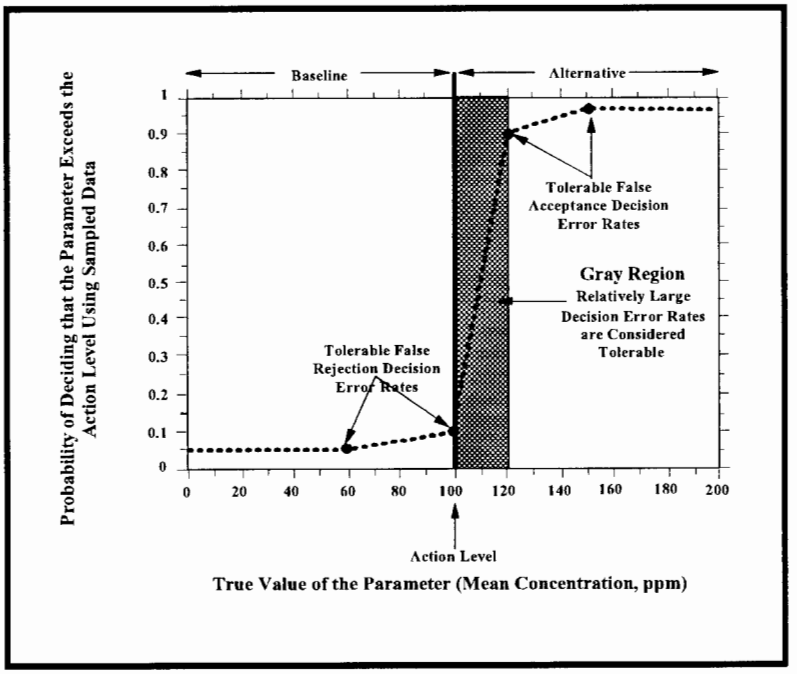


Figure 6-4. An Example of a Decision Performance Goal Diagram (Baseline Condition: Parameter is less than the Action Level)

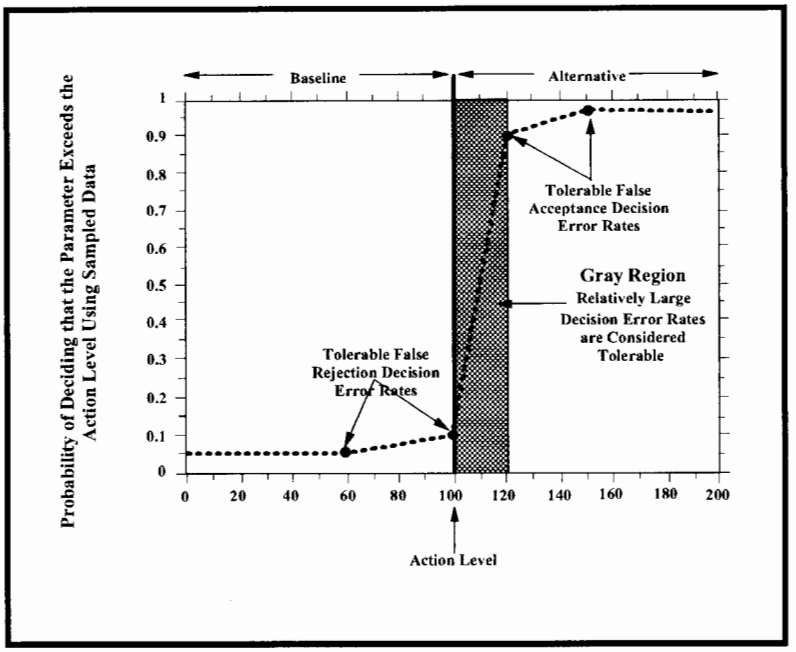


Figure 6-5. Decision Performance Goal Diagram for Example 1

6.4 Examples

Example 1. Making Decisions About Incinerator Fly Ash for RCRA Waste Disposal

How was the baseline condition set? *The baseline condition [i.e., the null hypothesis (H_0)] was established as "the waste is hazardous." The consequences of deciding that the waste was not hazardous when it truly was hazardous were that the incinerator company disposed of the waste in a sanitary landfill, possibly endangering human health and the environment. In this situation, the incinerator company could be held liable for future damages and environmental cleanup costs. Additionally, the consequences of this decision error were to compromise the reputation of the incinerator company, jeopardizing its future profitability. The planning team concluded that this decision error (false rejection) had the more severe consequences near the Action Level since the risk of jeopardizing human health outweighed the consequences of having to pay more for disposal.*

How was the gray region specified? *The gray region was designated as that area adjacent to the Action Level where the planning team considered that the consequences of a false acceptance decision error were minimal. The planning team specified a width of 0.25 mg/L for this gray region based on their preferences to guard against false acceptance decision errors at a concentration of 0.75 mg/L (the lower bound of the gray region).*

How were tolerable decision error limits set? *RCRA regulations specify a 5% decision error rate at the Action Level. Below the Action Level, the planning team set the maximum tolerable probability of making a false acceptance error at 20% when the true parameter was from 0.25 to 0.75 mg/L and 10% when it was below 0.25 mg/L. These limits were based on both experience and an economic analysis that showed that these decision error rates reasonably balanced the cost of sampling versus the consequence of sending clean ash to the RCRA facility.*

Example 2. Making Decisions About Urban Air Quality Compliance

How was the baseline condition set? *In most applications of the DQO Process, when, where, and how many samples to collect is not determined until Step 7. However, given that the monitoring network and sampling frequency were already established, the DQO Process in this case was conducted to establish the quality and quantity of data needed for making attainment decisions and to determine if the present network design achieved those quality and quantity specifications. As the planning team was most concerned about protecting public health, the baseline condition in this case was that the 98th percentile of daily concentrations was above 65 $\mu g/m^3$ (i.e., less than 98% of daily concentrations are below 65 $\mu g/m^3$). That is, the null hypothesis was set as the state of nature the planning team found evidence against, and, to protect public health, carefully guarded against the false rejection decision error of incorrectly rejecting the baseline condition.*

How was the gray region specified? *The gray region, in this case, was specified in terms of proportions. The planning team decided that the gray region should be from 0.98 to 0.995.*

How were tolerable decision error limits set? *The planning team determined that the tolerable false rejection decision error rate should be 10% or less. While lowering the tolerable bound on such error was desirable, the planning team, based on observed $PM_{2.5}$ daily concentration variability in other parts of the country, believed that significantly smaller false rejection error rates were unobtainable for all but the most extensive and costly network designs. The team also wished to protect against implementing* unnecessary and costly control strategies (i.e., incorrectly failing to reject the baseline condition), but was willing to tolerate a somewhat larger probability of making this false acceptance decision error. The planning team decided that the false acceptance decision error rate should be not larger than 30%. These are shown in Figure 6.6.

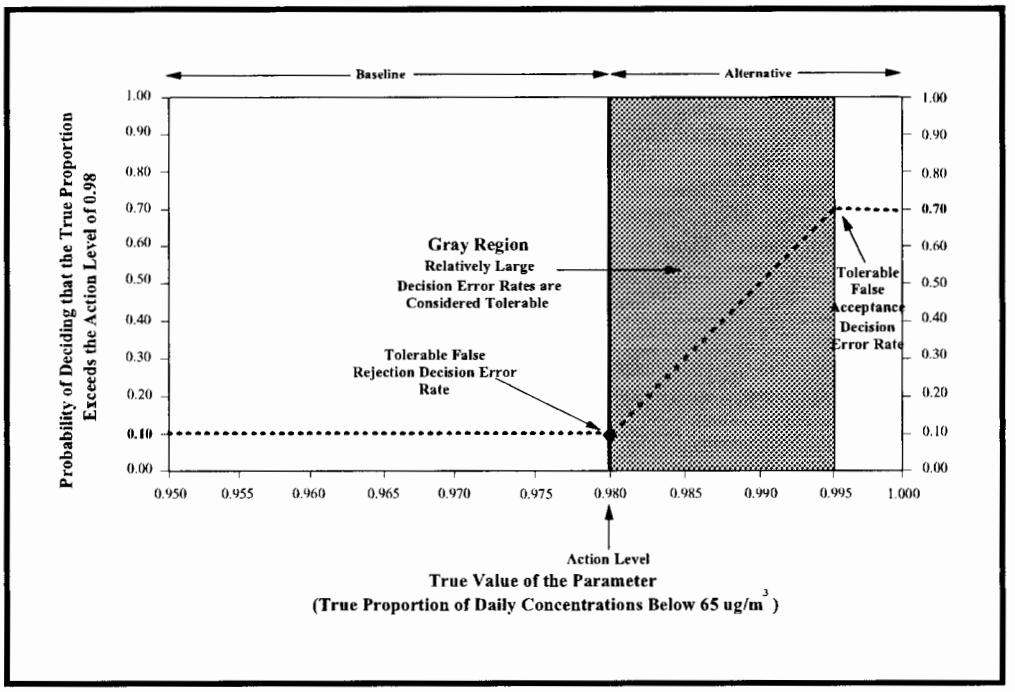


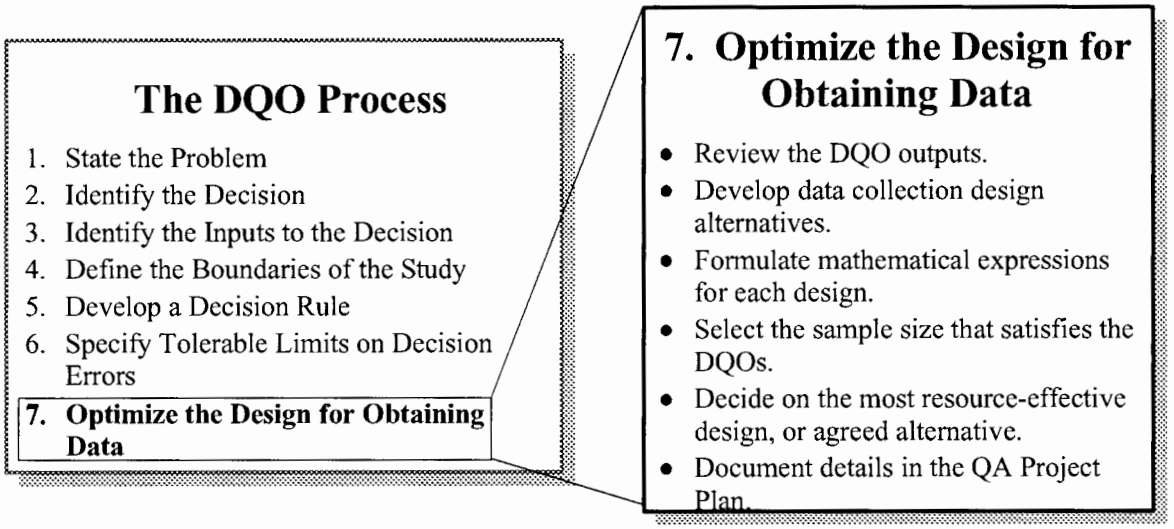
Figure 6-6. Decision Performance Goal Diagram for Example 2

Looking ahead to other DQO steps:

- The information developed in Step 6 is then translated into the requirements for a sampling plan in Step 7.

CHAPTER 7

STEP 7. OPTIMIZE THE DESIGN FOR OBTAINING DATA



After reading this chapter you should have a broad understanding of the steps that are needed to develop a sampling and analysis design to generate data that meet the Data Quality Objectives and Decision Performance Goals developed in Steps 1 through 6 of the DQO Process.

7.1 Background

This step builds on the previous steps where you have:

- identified members of the planning team, including decision makers;
- concisely described the problem;
- developed a conceptual model of the environmental problem to be investigated;
- identified the decision that needs to be made;
- determined the type of information required, the Action Level, and probable measurement methods;
- decided on the spatial/temporal boundaries of the decision and the scale of the decision making;
- decided on the theoretical "*if...then*" decision rule; and
- specified tolerable limits on decision errors.

The purpose of Step 7 is to develop a resource-effective sampling and analysis design for generating data that are expected to satisfy the DQOs and DPGs developed in Steps 1 through 6 of the DQO Process.

7.2 Activities

In this final step you should:

- review existing environmental data;
- evaluate operational decision rules;
- develop general data collection design alternatives;
- calculate the number of samples to be taken; and
- select the most resource-effective data collection design.

Why should you review existing environmental data? Review existing data in more detail if it appears that they can be used to support the data collection design (e.g., analyze the variability in existing data if they appear to provide good information about the variance for the new data). If no existing data are available, it may be cost-effective to conduct a limited field investigation to acquire preliminary estimates of variability for determining the number of samples. If existing data are going to be combined with new data to support the decision, then determine if there are any gaps that can be filled or deficiencies that might be mitigated by including appropriate features in the new data collection design. The existing data should also be reviewed for indications of analytical problems, such as detection limits, that may rule out using certain statistical techniques. Prior knowledge of the probability distribution (characteristics) exhibited by the data may also have an effect on the choice of statistical tests.

How do you evaluate operational decision rules? The theoretical decision rule you developed in Step 5 was based on the assumption that you knew the true value of the parameter of interest (e.g., the true mean or median). Since you will be using measurements made on samples to make your decision, an operational decision rule will be needed to replace the theoretical decision rule. This operational decision rule will most likely be in the form of a statistical hypothesis test which may involve some form of a statistical interval such as a confidence interval or tolerance interval. The design team should evaluate the possible operational decision rules and choose one that best matches the intent of the theoretical decision rule with the statistical assumptions. Each operational decision rule will have a different formula for determining the number of samples needed to meet your DPGs.

Some common statistical hypothesis tests and their sample size formulas are described in detail in *Guidance for Data Quality Assessment: Practical Methods for Data Analysis (EPA QA/G-9)*, (U.S. EPA, 1997a). Most tests applied to environmental data can be broadly classified as one-sample (single-site) tests or two-sample (two-site) tests. In one-sample cases, data from a site are compared with an absolute criterion such as a regulatory threshold or an Applicable or Relevant and Appropriate Requirement. In the two-sample cases, data from a site are compared

with data from another site or background (reference) area or from another time period at the same site. 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 (i.e., no difference).

How do you develop data collection design alternatives? A full explanation of the procedures for developing a data collection design is beyond the scope of this guidance document. This document provides a broad overview of the steps that need to be accomplished to reach a final sampling plan. This section provides a general description of the activities necessary to generate sampling design options and select the one that optimally satisfies the DPGs defined in Step 6. In addition, it contains information about how outputs from the previous six steps of the DQO Process are used in developing the most resource-effective sampling and analysis design.

The design team should develop alternative data collection and analysis designs based on the DQO outputs and other relevant information, such as historical patterns of contaminant deposition, estimates of variance, and technical characteristics of the contaminants and media. The most important element of this step is to reduce the total variability through judicious choice of a spatial and temporal sampling design and analytical measurement technique (see also Figure 6-1). If the total variability can be reduced to a value less than that specified in Step 6, the result will be either a reduction in decision error rates (given a fixed number of samples) or reduction in the number of samples (and, hence, resource expenditure) for a given set of decision error rates. In general, the more complex the sampling design, the lower the total variability of the sample will be.

Generally, the goal is to find cost-effective design alternatives that balance the number of samples and the measurement performance, given the feasible choices for spatial and temporal sample designs and measurement methods. In cases where there is relatively high spatial or temporal variability, it may be more cost-effective to use less expensive and less precise analytical methods so that a relatively large number of samples over space and time can be taken, thereby controlling the sampling design error component of total study error. In other cases, where the contaminant distribution over space and time is relatively homogeneous, or the Action Level is very near the method detection limit, it may be more cost-effective to use more expensive more precise and/or more sensitive analytical methods and collect fewer samples, thereby reducing the analytical measurement error component of total study error. These alternatives should, at a minimum, include the sample selection technique, the sample type, the number of samples, and the number of analyses per sample. To generate alternative designs, the planning team may vary the number and spatial/temporal locations of samples, the type of samples collected, the field sampling or analytical methods used, or the number of replicate analyses performed on samples. It should be remembered that the objective of the design is to estimate the parameter (mean, median, percentile) with as much precision as possible such that the DPGs can be achieved.

How do you calculate the number of samples that satisfy the DPGs for each design alternative and determine the cost for each design? You should use the formulas identified in

the previous activity to calculate the number of samples needed to meet the DPGs for each data collection design alternative. You should then determine the associated cost for each design alternative.

To assist the design team in their development of alternative designs and evaluation of costs for a few select sampling designs and operational decision rules, EPA has developed the software, *Data Quality Objectives Decision Error Feasibility Trials (DEFT) Software (EPA QA/G-4D),* (U.S. EPA, 1994b). DEFT is a personal computer software package developed to assist your planning team in evaluating whether the DQOs are feasible (i.e., can be achieved within resource constraints) before the development of the final data collection design is started. DEFT uses the outputs generated in Steps 1 through 6 of the DQO Process to evaluate several basic data collection designs and determines the associated cost. DEFT presents the results in the form of a Decision Performance Goal Diagram that overlays the desired Decision Performance Curve of the sampling design.

If the DQOs are not feasible or not achievable within resource constraints, the DEFT software allows you to relax some of the DQOs and DPGs until a feasible alternative is achieved. The software allows the user to change the action level, the baseline condition, the width of the gray region, the decision error rates, 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 and shows the corresponding Decision Performance Curve in the Decision Performance Goal Diagram.

How do you select the most resource-effective data collection design that satisfies all of the DQOs? You should evaluate the design options based on cost and ability to meet the DQO constraints and DPGs. The design that provides the best balance between cost (or expected cost) and ability to meet the DQOs, given the non-technical, economic, and health factors imposed on the project, is the most resource-effective (or the optimum design) .

The statistical concept of a power function is extremely useful in investigating the performance of alternative designs. The power function is the probability of rejecting the null hypothesis (H_0) when the null hypothesis is false (i.e., the alternative condition is true). If there was no error associated with a decision, the ideal power function would be 0 if H_0 were true, and 1 if H_0 were false. Since decisions are based on imperfect data, however, it is impossible to achieve this ideal power function. Instead, the power function will most likely yield values that are small when H_0 is true and large when H_0 is false. A performance curve is based on the graph of the power function.[2] The performance curve can be overlaid into the Decision Performance Goal Diagram to assess how well a test performs or to compare competing test. A design that produces a very steep performance curve is preferred over one that is relatively flat.

---

[2]In this guidance, the performance curve is based on either the power curve or the complement of the power curve. This ensures that the performance curve always rises from left to right.

One simple method to improve the power of the statistical design is the use of stratification to reduce the total variability in the data. Stratification is done by dividing the target population into strata that are relatively homogeneous. The planning team may have made an initial attempt at this in Step 4, Define the Boundaries of the Study. The strata may be physically based (areas proximal to an incinerator, septic tanks, receptor wells, underground storage tanks) or based on other factors (potential exposure, activity patterns, residences, ecological habitats, agricultural sectors, historical or future use).

The advantages of stratification are:

- reducing the complexity of the problem by dividing it into manageable segments;
- reducing the variability in subsets; and
- improving the efficiency of sampling.

Disadvantages of stratification include:

- difficulty in determining the basis for selecting strata (prior estimates of variability, estimates of strata area may be needed);
- overstratifying may require more samples so increasing costs; and
- stratifying areas that are not approximately homogeneous may result in developing a design for collecting data that is inefficient or does not accurately reflect the characteristics of the population.

If none of the data collection designs satisfies all of the DQOs and DPGs within the resource constraints of the project, the planning team will need to review the outputs from the entire DQO Process and alter one or more of the steps. Examples of adjustments that could be made are:

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

For other sampling designs and/or operational decision rules, it will be necessary for the design team to evaluate the design alternatives by other methods (perhaps computer simulation) and possibly involve a statistical expert on sampling design and analysis.

Application of the DQO Process to remediation problems and integration with geostratistical approaches to the analysis of soil contamination scenarios may be found in Myers (1997). Once the final data collection design has been selected, it is important to ensure the design and operational decision rule are properly documented. This improves efficiency and

effectiveness of later stages of the data 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 and operational decision rule are 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, Sampling and Analysis Plan, QA Project Plan or other required document. Design elements that should be documented include:

- number of samples,
- 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),
- physical sample (i.e., the amount of material to be collected for each sample),
- sample support (i.e., the area, volume, or quantity that each individual sample represents),
- sample locations (surface coordinates and depth) and how locations were selected,
- timing issues for sample collection, handling, and analysis,
- analytical methods (or performance-based measurement standards), and
- statistical sampling scheme.

Note that proper documentation of the model, operational decision rule, and associated assumptions used for collecting and statistically analyzing data is essential to maintain the overall validity of the study in the face of unavoidable deviations from the original design. Additionally, the documentation will serve as a valuable resource for Data Quality Assessment (DQA) activities after the data have actually been collected and the subsequent decision making process has been completed.

7.3 Outputs

The outputs from this step are the full documentation of the final sampling design and discussion of the key assumptions supporting the sampling design.

7.4 Examples

The examples presented here represent the initial final output of the DQO Process.

Example 1. Making Decisions About Incinerator Fly Ash for RCRA Waste Disposal

What was the selected sampling design? *The planning team's statistician performed an initial cost/benefit analysis that indicated a composite sample design was the best sampling option to use to determine whether a container of ash should be sent to a RCRA landfill or to a municipal landfill. Eight composite samples, each consisting of*

eight grab samples, were taken from each container; and two subsamples from each composite were sent to the laboratory for analysis. To form the composite samples, the containers were divided into eight equal size areas and grab samples were taken randomly within each area and composited. Each grab sample was a core that was extracted, then mixed together to form the composite sample. From this composite sample, two subsamples were sent to the laboratory for analysis.

What were the key assumptions supporting the selected design? *The cost of this design was based on the cost of collecting ($10) and analyzing ($150) a sample. Eight grab samples were collected for each composite sample, for a sampling cost of $80; two subsamples were analyzed from each composite sample for a cost of $300. Therefore, each composite sample cost $380. The total cost of collecting and analyzing the eight composite samples in one container was eight times the cost of one composite, for a total of $3,040. The assumption that composite measurements were normally distributed was made. This assumption was evaluated after the measurements were obtained. If the assumption was not viable, then the planning team would recommend that additional grab samples per composite be taken, or that a revised compositing process be used to achieve normally distributed data. Based on the pilot study, the incineration company determined that each load of waste fly ash was fairly homogeneous and estimated the standard deviation in the concentration of cadmium among grab samples within loads of ash to be 0.6 mg/L. It was assumed that the variability among sub-samples within a composite sample was negligible. Data from the subsamples was used to test this assumption and to collect additional subsamples, if necessary.*

Example 2. Making Decisions about Urban Air Quality Compliance

What was the selected sampling design? *Information from Step 6 indicated that sampling everyday, regardless of the false rejection decision error rate tolerated, was probably an inefficient use of resources and was unnecessary. This conclusion was reached because sampling daily resulted in false acceptance decision error rates that were far below those required in Step 6. In contrast, 1-in-6-day or 1-in-3-day sampling could not satisfy the false acceptance decision error rate of 30% when the rather restrictive constraint of a 1% false rejection decision error rate was used. The current sampling scheme (1 in 3 days) performed at a satisfactory level as long as the false rejection decision error rate allowed was in the range of 5% to 10%. If the planning team decided that up to 10% false rejection decision error truly could be tolerated, then information in Table 7-1 indicated it was possible to reduce sampling frequency from the current rate to 1-in-6-day sampling, thereby reducing costs while maintaining an acceptable false acceptance decision error rate around 23%.*

What were the key assumptions supporting the selected design? *The monitoring network was already in place, so the goal at this stage was to determine the performance of the existing design, and to change the design, if needed, to achieve better performance. Information in Table 7-1 showed the design performance (false acceptance decision error rate) as a function of different false rejection error rate allowances and alternative sampling frequencies (sample sizes) over a 3-year period of data collection. In general, data in Table 7-1 indicated that the false acceptance*

decision error rate decreased when a higher false rejection decision error rate was tolerated. Similarly, false acceptance decision error rates decreased when sampling intensity was increased from one-in-six-days sampling to every-day sampling.

Table 7-1. False Acceptance Decision Error Rates and Alternative Sampling Frequencies

		Sampling Frequency At Each Of Three Monitors		
		1 in 6 Days	1 in 3 Days (Current)	Every Day
Tolerable False Rejection Decision Error Rates	1%	>50%	>50%	1%
	5%	>50%	28%	<1%
	10%	23%	11%	<1%

CHAPTER 8

BEYOND THE DATA QUALITY OBJECTIVES PROCESS

After reading this chapter you should understand the kinds of information that will be necessary to develop a QA Project Plan and the role of Data Quality Assessment.

A project's life cycle consists of three principal phases: planning, implementation, and assessment (described in Chapter 0 as the project tier of EPA's Quality System). Quality assurance activities that are associated with each of these phases are illustrated in Figure 0-1. Systematic planning (e.g., the DQO Process) and developing the sampling design comprises the planning phase; the actual data collection process is the implementation phase; and an evaluation (Data Quality Assessment) that the collected data met the performance criteria specified in the DQOs is the final phase of a project.

8.1 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. Systematic planning and the DQO Process are the foundation for the planning stage and lead to a sampling design, the generation of appropriate data quality indicators, and standard operating procedures, which are all finally documented in the Agency's mandatory QA Project Plan or similar document.

Environmental data for EPA programs may not be collected without having an approved QA Project Plan in place (EPA Order 5360.1 A2). The mandatory QA Project Plan (EPA, 1998) documents four main groups – project management, data generation and acquisition, assessment/oversight, and data validation and usability (shown in Table 8-1).

Group A – Project Management
These elements address project management, project history and objectives, and roles and responsibilities of the participants. These elements help ensure that project goals are clearly stated, that all participants understand the project goals and approach, and that the planning process is documented.

Group B – Data Generation and Acquisition
These elements cover all aspects of the project design and implementation (including the key parameters to be estimated, the number and type of samples expected, and a description of where, when, and how samples will be collected). They ensure that appropriate methods for sampling, analysis, data handling, and QC activities are employed and documented.

Table 8-1. Elements of a Quality Assurance Project Plan

QA Project Plan Elements	
A. Project Management	
A1 Title and Approval Sheet	A6 Project/Task Description
A2 Table of Contents	A7 Quality Objectives and Criteria
A3 Distribution List	A8 Special Training /Certification
A4 Project/Task Organization	A9 Documents and Records
A5 Problem Definition/Background	
B. Data Generation and Acquisition	
B1 Sampling Process Design (Experimental Design)	B7 Instrument/Equipment Calibration and Frequency
B2 Sampling Methods	B8 Inspection/Acceptance of Supplies and Consumables
B3 Sample Handling and Custody	B9 Nondirect Measurements
B4 Analytical Methods	B10Data Management
B5 Quality Control	
B6 Instrument/Equipment Testing, Inspection, and Maintenance	
C. Assessment and 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

Group C – Assessment and Oversight
These elements address activities for assessing the effectiveness of project implementation and associated QA and QC requirements; they help to ensure that the QA Project Plan is implemented as prescribed.

Group D – Data Validation and Usability
These elements address QA activities that occur after data collection or generation is complete; they help to ensure that data meet the specified criteria.

Additional information on the preparation of QA Project Plans is provided in EPA's guidance document, *Guidance on Quality Assurance Project Plans (EPA QA/G-5)* (U.S. EPA, 1998).

8.2 Implementation

During the implementation phase of the project, data are collected and samples are analyzed according to the specifications of the QA Project Plan or the Sampling and Analysis Plan depending on specific program requirements. These provide detailed specific objectives, QA and QC specifications, and procedures for conducting a successful field investigation that is intended to produce data of the quality needed to satisfy the performance criteria. QA and QC activities (e.g., 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.

8.3 Assessment

During the final phase (assessment) of a project, data are verified and validated in accordance with the QA Project Plan, and DQA is performed to determine if the performance criteria have been satisfied.

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; you really should 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. DQA is the scientific and statistical process that determines whether environmental data are of the right type, quality, and quantity to support a specific decision. DQA consists of five steps that parallel the activities of a statistician analyzing a data set; and include the use of statistical and graphical tools that nonstatisticians can apply to data sets (see Figure 8-1).

DQA involves the application of statistical tools to determine whether the data are of appropriate quality to support the decision with acceptable confidence. 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 explain statistical nomenclature and avoid the use of statistical jargon whenever possible. For more information on Data Quality Assessment, see EPA's guidance document, *Guidance for Data Quality Assessment: Practical Methods for Data Analysis (EPA QA/G-9)*, (U.S. EPA, 1997a) and the associated software Data Quality Assessment Statistical Toolbox (DataQUEST) (*EPA QA/G-9D*), (U.S. EPA, 1997b).

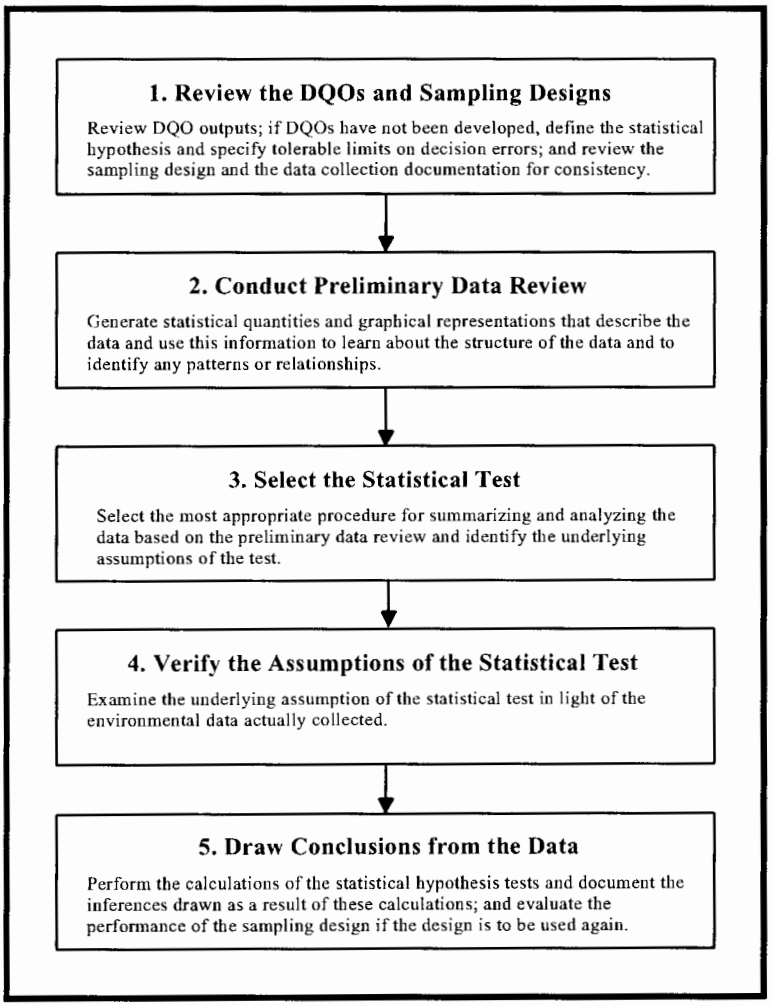


Figure 8-1. Data Quality Assessment Process

APPENDIX A

DERIVATION OF SAMPLE SIZE FORMULA FOR TESTING MEAN OF NORMAL DISTRIBUTION VERSUS AN ACTION LEVEL

This appendix presents a mathematical derivation of the sample size formula used in the DQO Example 1.

Let $X_1, X_2, \ldots, X_n$ denote a random sample from a normal distribution with unknown mean μ and known standard deviation σ. The decision maker wishes to test the null hypothesis H_0: $\mu = AL$ versus the alternative H_A: $\mu > AL$, where AL, the action level, is some prescribed constant; the false positive (Type I) error rate is α (i.e., probability of rejecting H_0 when $\mu = AL$ is α); and for some fixed constant $U > AL$ (where U is the other bound of the gray region), the false negative (Type II) error rate is β (i.e., probability of rejecting H_0 when $\mu = U$ is $1-\beta$). Let $\bar{X}$ denote the sample mean of the Xs. It will have a normal distribution with mean μ and variance σ^2/n. Hence the random variable Z, defined by

$$Z = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \quad , \qquad \text{(A - 1)}$$

will have a standard normal distribution (mean 0, variance 1). Let z_p denote the pth percentile of the standard normal distribution (available in most statistics books). Recall that the symmetry of the standard normal distribution implies that $z_p = -z_{1-p}$.

Case 1: Standard Deviation Known

The test of H_0 versus H_A is performed by calculating the test statistic.

$$T = \frac{(\bar{X} - AL)\sqrt{n}}{\sigma}. \qquad \text{(A - 2)}$$

If $T > z_{1-\alpha}$, the null hypothesis is rejected.

Note that

$$T = \frac{[(\bar{X} - \mu) + (\mu - AL)]\sqrt{n}}{\sigma} = Z + \delta(\mu) \qquad \text{(A - 3)}$$

where

$$\delta(\mu) = \frac{(\mu - AL)\sqrt{n}}{\sigma}. \qquad \text{(A - 4)}$$

Thus T has a normal distribution with mean $\delta(\mu)$ and variance 1, and, in particular, $\delta(AL) = 0$. Hence the Type I error rate is

$$Pr[rejecting\ H_0 | H_0] = Pr[T > z_{1-\alpha} | \mu = AL] = Pr[Z + \delta(AL) > z_{1-\alpha}] = \alpha \qquad \text{(A - 5)}$$

Achieving the desired power 1-β when μ = U requires that

$$Pr[reject\ H_0 | \mu = U] = 1 - \beta.$$

Therefore,

$$Pr[T \le z_{1-\alpha} | \mu = U] = Pr[Z + \delta(U) \le z_{1-\alpha}] = Pr[Z \le z_{1-\alpha} - \delta(U)] = \beta. \qquad \text{(A - 6)}$$

This implies

$$z_{1-\alpha} - \delta(U) = z_{\beta},$$

or

$$z_{1-\alpha} - \frac{(U - AL)\sqrt{n}}{\sigma} = -z_{1-\beta}.$$

Let Δ = U-AL, then rearrange terms to obtain

$$(z_{1-\alpha} - z_{1-\beta})\sigma = \Delta\sqrt{n},$$

or

$$n = \frac{(z_{1-\alpha} - z_{1-\beta})^2 \sigma^2}{\Delta^2}. \qquad \text{(A - 7)}$$

Case 2: Standard Deviation Unknown

If the standard deviation σ is unknown, then a test statistic such as Equation A - 2 is used except that σ is replaced by S, an estimate of the standard deviation calculated from the observed Xs. Such a statistic has a noncentral t distribution rather than a normal distribution, and the n computed by the above formula will be too small, although for large n (say n>40), the approximation is good. The particular noncentral t distribution involved in the calculation depends on the sample size n. Thus, determining the

exact minimum n that will satisfy the Type I and Type II error rate conditions requires an iterative approach in which the noncentral t probabilities are calculated for various n values until the desired properties are achieved. With the aid of a computer routine for calculating such probabilities, this is not difficult; however, a simple and direct approach for approximating n is available. This approach, whose derivation is described in the paragraphs below, leads to the following approximate but very accurate formula for n:

$$n \approx \frac{(z_{1-\alpha} + z_{1-\beta})^2 \sigma^2}{\delta^2} + \frac{1}{2} z_{1-\alpha}^2. \qquad \text{(A - 8)}$$

In practice, since σ is unknown, a prior estimate of it must be used in Equation A - 8.

The approach is based on the assumption that, for a given constant k, the statistic $\bar{X} - kS$ is approximately normal with mean μ-kσ and variance $(\sigma^2/n)(1+k^2/2)$ (Guenther, 1977 and 1981).

The classical t-test rejects H_0 when, $T = [(\bar{X} - AL)/S\sqrt{n})] > D$ where the critical value D is chosen to achieve the desired Type I error rate α. The inequality can be rearranged as $\bar{X} - kS > AL$, *where* $k = D\sqrt{n}$. Subtracting the mean (assuming H_0) and dividing by the standard deviation of $\bar{X} - kS$ on both sides of the inequality leads to

$$\frac{\bar{X} - kS - (AL - k\sigma)}{(\sigma/\sqrt{n})\sqrt{1 + k^2/2}} > \frac{AL - (AL - k\sigma)}{(\sigma/\sqrt{n})\sqrt{1 + k^2/2}} = \frac{k\sqrt{n}}{\sqrt{1 + k^2/2}}. \qquad \text{(A - 9)}$$

By the distributional assumption on $\bar{X} - kS$, the left side of Equation A - 9 is approximately standard normal when μ = AL, and the condition that the Type I error rate is α becomes

$$Pr[Z > k\sqrt{n}/\sqrt{1 + k^2/2}] \approx \alpha, \qquad \text{(A - 10)}$$

$$i.e., \quad z_{1-\alpha} \approx k\sqrt{n}/\sqrt{1 + k^2/2}. \qquad \text{(A - 11)}$$

One can show that Equation A - 11 is equivalent to

$$1/[1 - k^2/2] \approx 1 - z_{1-\alpha}^2/2n. \qquad \text{(A - 12)}$$

The condition that the Type II error rate is β (or that power is 1-β) when μ = U means that the event of incorrectly accepting H_0 given $\bar{X} - kS \le AL$ should have probability β. Subtracting the mean (U - kσ) and dividing by the standard deviation of $\bar{X} - kS$ on both sides of this inequality yields

$$\frac{\bar{X} - kS - (U - k\sigma)}{(\sigma/\sqrt{n})\sqrt{1 - k^2/2}} \geq \frac{AL - (U - k\sigma)}{(\sigma/\sqrt{n})\sqrt{1 - k^2/2}}. \qquad \text{(A - 13)}$$

Again, the left side is approximately standard normal and the Type II error rate condition becomes

$$Pr\{Z \leq [AL - (U - k\sigma)]/[(\sigma/\sqrt{n})/\sqrt{1 - k^2/2}]\} \leq \beta,$$

which implies

$$-z_{1-\beta} \leq z_{\beta} \leq \frac{(AL - U) - k\sigma}{(\sigma/\sqrt{n})\sqrt{1 - k^2/2}}. \qquad \text{(A - 14)}$$

Subtracting Equation A - 14 from Equation A - 11 yields

$$z_{1-\alpha} - z_{1-\beta} \geq \frac{(U - AL)}{(\sigma/\sqrt{n})\sqrt{1 - k^2/2}}, \qquad \text{(A - 15)}$$

or

$$\frac{z_{1-\alpha} - z_{1-\beta}}{(U - AL)} \geq \frac{\sqrt{n}}{\sqrt{1 - k^2/2}}. \qquad \text{(A - 16)}$$

Substituting Equation A - 12 into the denominator on the right side of Equation A - 16 yields

$$\frac{z_{1-\alpha} - z_{1-\beta}}{(U - AL)} \geq \sqrt{n}\sqrt{1 - z_{1-\alpha}^2/2n}. \qquad \text{(A - 17)}$$

Squaring both sides of Equation A - 17 and solving for n yields Equation A - 8.

References

Guenther, William C. 1977. *Sampling Inspection in Statistical Quality Control.* Griffin's Statistical Monographs and Courses, No. 37, London: Charles Griffin.

Guenther, William C. 1981. Sample size formulas for normal theory T test. *The American Statistician* 35, 4.

APPENDIX B

BIBLIOGRAPHY

ANSI/ASQC E4-1994, *Specifications and Guidelines for Environmental Data Collection and Environmental Technology Program.* 1994. American Society for Quality, Milwaukee, WI.

40 CFR 50, Code of Federal Regulations, "National Ambient Air Quality Standards for Particulate Matter; Availability of Supplemental Information and Request for Comments, Final Rule." Washington, DC.

40 CFR 261, Code of Federal Regulations, Appendix II, "Identification and Listing of Hazardous Waste; Method 1311 Toxicity Characteristic Leaching Procedure (TCLP)." Washington, DC.

40 CFR 745, Code of Federal Regulations, "Lead; Identification of Dangerous Levels of Lead; Proposed Rule." Washington, DC.

ASTM E1613-99, "Standard Test Method for Determination of Lead by Inductively Coupled Plasma Atomic Emission Spectrometry (ICP-AES), Flame Atomic Absorption Spectrometry (FAAS), or Graphite Furnace Atomic Absorption Spectrometry (GFAAS) Techniques," American Society for Testing and Materials, 1999.

ASTM E1644-98, "Standard Practice for Hot Plate Digestion of Dust Wipe Samples for the Determination of Lead," American Society for Testing and Materials, 1998.

ASTM E1728-99, "Standard Practice for Field Collection of Settled Dust Samples Using Wipe Sampling Methods for Lead Determination by Atomic Spectrometry Techniques," American Society for Testing and Materials, 1999.

Gilbert, R.O. *Statistical Methods for Environmental Pollution Monitoring.* 1987. John Wiley, New York.

Grumley, Thomas. 1994. *Institutionalizing the Data Quality Objectives Process for EM's Environmental Data Collection Activities.* Department of Energy, Washington, DC.

Harbinger Communications. 1996. *The Harbinger File: A Directory of Citizen Groups, Government Agencies, and Environmental Education Programs Concerned with California Environmental Issues.* Santa Cruz, CA.

HUD (Department of Housing and Urban Development). 1995. *Guidelines for the Evaluation and Control of Lead-based Paint Hazards in Housing*. Office of Lead-based Paint Abatement and Poisoning Prevention, Washington, DC.

Myers, J.C. *Geostatistical Error Management: Quantifying Uncertainty for Environmental Sampling and Mapping*. 1997. Van Nostrand Reinhold. New York.

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

U.S. EPA (Environmental Protection Agency). 1994a. *Guidance for the Data Quality Objectives Process (EPA QA/G-4)*. EPA/600/R-96/055. Washington, DC.

U.S. EPA (Environmental Protection Agency). 1994b. *Data Quality Objectives Decision Error Feasibility Trials (DEFT) Software (EPA QA/G-4D)*. EPA/600/R-96/056. Washington, DC.

U.S. EPA (Environmental Protection Agency). 1996. *Soil Screening Guidance: User's Guide*. EPA/540/R-96/0180. Washington, DC.

U.S. EPA (Environmental Protection Agency). 1997a. *Guidance for Data Quality Assessment: Practical Methods for Data Analysis (EPA QA/G-9)*. EPA/600/R-96/084. Washington, DC.

U.S. EPA (Environmental Protection Agency). 1997b. *Data Quality Assessment Statistical Toolbox (DataQUEST) (EPA QA/G-9D)*. EPA/600/R-96/085. Washington, DC.

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

U.S. EPA (Environmental Protection Agency). 1999. *Data Quality Objectives for Hazardous Waste Site Investigations (EPA QA/G-4HW)*. Washington, DC.

U.S. EPA (Environmental Protection Agency). 2000a. *EPA Order 5360.1 A2: Policy and Program Requirements for the Mandatory Agency-wide Quality System..* Washington, DC.

U.S. EPA (Environmental Protection Agency). 2000b. *EPA Order 5360: The EPA Quality Manual for Environmental Programs*. Washington, DC.

APPENDIX C

Data Quality Objectives; Household Dust Lead Hazard Assessment

This example concerns the use of the median in planning for environmental decision making. The example is presented in a continuous format to show the seven-step DQO Process in its entirety.

0. Background

The adverse health effects resulting from exposure to lead hazards (paint, dust, and soil) have received increasing attention because chronic exposure to low levels of lead can cause impairment of the central nervous system, mental retardation, and behavioral disorders. Young children (below the age of six) are at a particularly high risk for these adverse effects. Concern about the exposure to lead hazards in residential housing has led federal agencies, including the EPA and Department of Housing and Urban Development, to develop programs to evaluate, and ultimately control, lead hazards in housing.

A critical pathway for exposure to lead by a child is through the ingestion of household dust because dust collects on hands, toys, and food and is easily transferred by hand-to-mouth activities. As a result of the concern about the dust-to-mouth pathway, an important component of risk assessment is dust sampling. Dust sampling offers a way of characterizing dust lead levels at a property and determining if intervention is warranted. One of the preferred methods for sampling residential dust is using baby wipes to wipe a specified surface area. A single area may be sampled using an individual wipe; or multiple areas of a room may be sampled with individual wipes, and the individual wipes combined, or composited, then submitted to the laboratory as a single sample (40 CFR 745). The distribution of dust lead levels is such that normality cannot be assumed and a 50th percentile (the median) is the appropriate risk assessment level. This example demonstrates use of the median (i.e., 50th percentile) as the primary population parameter of concern

1. State the Problem

How were the planning team members selected? The planning team included the property owners, a certified risk assessor (to collect and handle dust samples and serve as a liaison with the laboratory), a statistician, and a quality assurance specialist. The decision makers were the property owners.

How was the problem described and a conceptual model of the potential hazard developed? The problem was described as evaluating potential hazards associated with lead in dust in a single-family residence because other residences in the neighborhood had shown levels of lead in dust that might pose potential hazards.

The conceptual model described a single-family residence in a neighborhood where hazardous levels of lead had been detected in other residences. Interior sources of lead in dust were identified as lead-based paint on doors, walls, and trim, which deteriorated to form, or attach to, dust particles. Exterior sources included lead in exterior painted surfaces that had deteriorated and leached into the dripline soil, or lead deposited from gasoline combustion fumes that accumulated in soil. In these cases, soil could be tracked into the house, and collected as dust on floors, window sills, toys, etc. Because this dust could be easily ingested through hand-to-mouth activities, dust was considered to be a significant exposure route. Levels of lead in floor dust were to be used as an indicator of the potential hazard.

What were the available resources and relevant deadlines? The property owners were willing to commit up to $1,000 for the study. To minimize inconvenience to the family, all sampling would be conducted during one calendar day.

2. Identify the Decision

What was the Decision Statement? The decision statement was determining if there were significant levels of lead in floor dust at the residence.

What were the alternative actions? If there were significant levels of lead in floor dust at the residence, the team planned follow-up testing to determine whether immediately dangerous contamination exists and the location of the contamination in the property. If not, the team decided there was not a potential lead hazard, and testing was discontinued.

3. Identify the Inputs to the Decision

Identify the kind of information. The assessment of a dust lead hazard was evaluated by measuring dust lead loadings by individual dust wipe sampling.

Identify the source of information. The EPA proposed standard stated that if dust lead levels were above 50 $\mu g/ft^2$ on bare floors, a lead health hazard was possible and follow-up testing and/or intervention should be undertaken (40 CFR 745).

What sampling and analytical methods were appropriate? Wipe samples were collected according to ASTM standard practice E1728. These samples were digested in accordance with ASTM standard practice E1644 and the sample extracts were chemically analyzed by ASTM standard test method E1613. The results of these analyses provided information on lead loading (i.e., µg of lead per square foot of wipe area) for each dust sample. The detection limit was well below the Action Level.

4. Define the Boundaries of the Study

What population was sampled? Dust contained in 1 ft^2 area of floors of the residence was sampled and sent to a laboratory for analysis.

What were the spatial boundaries? The spatial boundaries of the study area were defined as all floor areas within the dwelling that were reasonably accessible to young children who lived at, or visited, the property.

What was an appropriate time frame for sampling? The test results were considered to appropriately characterize the current and future hazards. It was possible that lead contained in soil could be tracked into the residence and collect on surfaces, but no significant airborne sources of lead deposition were known in the region. The dust was not expected to be transported away from the property; therefore, provided the exterior paint was maintained in intact condition, lead concentrations measured in the dust were not expected to change significantly over time.

What were the practical constraints for collecting data? Permission from the residents was required before risk assessors could enter the residence to collect dust wipe samples. Sampling was completed within 1 calendar day to minimize the inconvenience to the residents.

What was the scale of the decision making? The decision unit was the interior floor surface (approximately 1,700 ft^2) of the residence at the time of sampling and in the near future.

5. Develop a Decision Rule

What was the decision rule and Action Level? From 40 CFR 745, the median was selected as the appropriate parameter to characterize the population under study. The median dust lead loading was defined to be that level, measured in $\mu g/ft^2$, above and below which 50% of all possible dust lead loadings at the property were expected to fall. If the true median dust loading in the residence was greater than 50 $\mu g/ft^2$, then the planning team required followup testing. Otherwise, they decided that a dust lead hazard was not present and discontinued testing.

6. Specify Tolerable Limits on Decision Errors

How was the baseline condition set? The baseline condition adopted by the property owners was that the true median dust lead loading was above the EPA hazard level of 50 $\mu g/ft^2$ due to the seriousness of a potential hazard. The planning team decided that the most serious decision error would be to decide that the true median dust lead loading was below the EPA hazard level of 50 $\mu g/ft^2$, when in truth the median dust lead loading was above the hazard level. This incorrect decision would result in significant exposure to dust lead and potential adverse health effects.

How was the gray region specified? The edge of the gray region was designated by considering that a false acceptance decision error would result in the unnecessary expenditure of scarce resources for follow-up testing and/or intervention associated with a presumed hazard that did not exist. The planning team decided that this decision error should be adequately controlled for true dust lead loadings of 40 $\mu g/ft^2$ and below.

How were tolerable decision error limits set? Since human exposure to lead dust hazards causes serious health effects, the planning team decided to limit the false rejection error rate to 5%. This meant that if this dwelling's true median dust lead loading was greater than 50 $\mu g/ft^2$, the baseline condition would be correctly rejected 19 out of 20 times. The false acceptance decision, which would result in unnecessary use of testing and intervention resources, was allowed to occur more frequently (i.e., 20% of the time when the true dust-lead loading is 40 $\mu g/ft^2$ or less). These are shown in Figure C-1.

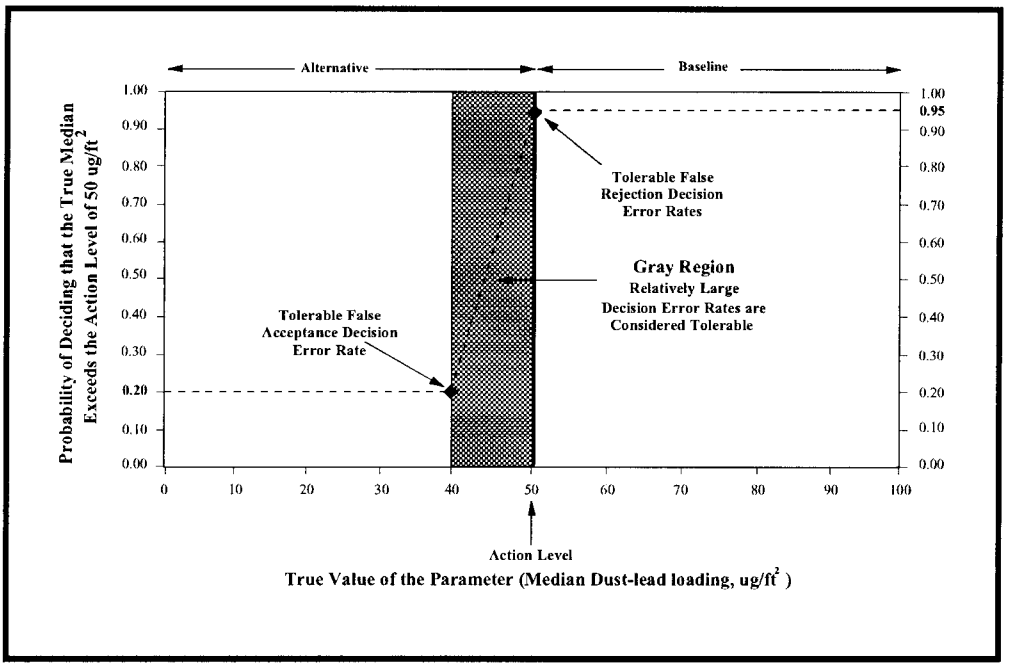


Figure C-1. Decision Performance Goal Diagram for Dust Lead Loading

7. Optimize the Design for Obtaining Data

What was the selected sampling design? The planning team determined that the cost of sending a certified risk assessor to the property for collecting and handling dust wipe samples was about $400. Also, an NLLAP-recognized laboratory was selected to analyze the collected wipe samples at a cost of $10 per sample. Thus, a maximum of 60 samples could be obtained within the study's cost constraint of $1,000. From Step 6 the initial gray region lower bound for the study was set at 40 $\mu g/ft^2$, but, the team found that this requirement could not be met given the specified decision errors (i.e., false rejection rate of 5% and false acceptance rate of 20%),

assumed standard deviation (of the natural logarithms), range, and cost constraints of the study (i.e., a maximum of 60 samples). The planning team decided they were unwilling to relax the decision error rate requirements and elected to expand the width of the gray region from the original 40 to 50 $\mu g/ft^2$ to the less restrictive range of 35 to 50 $\mu g/ft^2$. Further, the planning team decided that a standard deviation (of the natural logarithms) value of □=1.0 was probably more realistic than the more conservative estimate of □=1.5.

The planning team used the upper variability bound to develop Table C-1 which presented statistical sample size requirements across various assumed dust lead loading standard deviations (of the natural logarithms) and various lower bounds of the gray region. This table indicated that sample size requirements increased rather dramatically as variability increased and/or as the gray region was made more narrow.

Therefore, based on Table C-1, the planning team decided that a total of 50 samples should be collected by a certified risk assessor (all within 1 calendar day) using simple random sampling throughout the residence. Samples were sent to the selected NLLAP-recognized laboratory for analysis. The total study cost was approximately $900 to the property owners.

What were the key assumptions supporting the selected design? The dust lead loading data was assumed to be log-normally distributed. The geometric mean was computed using the data because the true median and true geometric mean are the same when log-normality is assumed. The true variability in dust lead loadings was not known, but past data was used to estimate a reasonable upper bound on variability.

Table C-1. Number of Samples Required for Determining If the True Median Dust Lead Loading is Above the Standard

Gray Region ($\mu g/ft^2$)	Standard Deviation of Natural Logarithms		
	□=0.5	**□=1.0**	**□=1.5**
20-50	6	9	13
25-50	8	15	21
30-50	14	26	37
35-50	26	50	75
40-50	64	126	188
45-50	280	559	837

APPENDIX D

GLOSSARY OF TERMS

acceptance criteria - specific limits placed on characteristics of an item, process, or service defined in requirements documents.

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 technology 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 rue 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.

alternative hypothesis - see alternative condition.

baseline condition - a tentative assumption to be proven either true or false. When hypothesis testing is applied to site assessment decision, 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 of volume (spatial boundary) and the time period (temporal boundary) to which a decision will apply.

confidence interval - the numerical interval constructed around a point estimate of a population parameter, combined with a probability statement (the confidence coefficient) linking to the population's true parameter value. If the same confidence interval construction technique and assumptions are used to calculate future intervals, they will include the unknown population parameter with the same specified probability.

data collection design - see sampling design.

data quality assessment (DQA) - a statistical and scientific evaluation of the data set to determine the validity and performance of the data collection design and statistical test, and to determine the adequacy of the data set for its intended use.

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

decision performance curve - a graphical representation of the quality of a decision process. In statistical terms it is know as a power curve (or a reverse power curve depending on the hypotheses being tested).

decision performance goal diagram (DPGD) - a graphical representation of the tolerable risks of decision errors. It is used in conjunction with the decision performance curve.

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

detection limit (DL) - a measure of the capability of an analytical method of distinguish samples that do not contain a specific analyte from sample that contain low concentrations of the analyte; the lower concentration or among of the target analyte that can be determine to be different from zero by a single measurement at a stated level of probability. DLs are analyte- and matrix-specific and may be laboratory dependent.

distribution - (1) the appointment of an environmental contaminant at a point over time, over an area, or within a volume; (2) a probability function (density function, mass function, or distribution function) used to describe a set of observations (statistical sample) or a population from which the observations are generated.

environmental conditions - the description of a physical medium (e.g., air, water, soil, sediment) or a biological system expressed in terms of its physical, chemical, radiological, or biological characteristics.

environmental data -any measurements or information that describe environmental processes, location, or conditions; ecological or health effects and consequences; or the performance of environmental technology. For EPA, environmental data include information collected directly from measurements, produced from models, and compiled from other sources such as data bases or the literature.

environmental processes - manufactured or natural processes that produce discharges to, or that impact, the ambient environment.

environmental programs - work or activities involving the environment, including but not limited to: characterization of environmental processes and conditions; environmental monitoring; environmental research and development; and the design, construction, and operation of environmental technologies; and laboratory operations on environmental samples.

environmental technology - an all-inclusive term used to describe pollution control devices and systems, waste treatment processes and storage facilities, and site remediation technologies and their components that may be utilized to remove pollutants or contaminants from, or to prevent them from entering, the environment. Examples include wet scrubbers (air), soil washing (soils), granulated activated carbon unit (water), and filtration (air, water). Usually, this term applies to hardware-based systems; however, it can also apply to methods or techniques used for pollution prevention, pollutant reduction, or containment of contamination to prevent further movement of the contaminants, such as capping, solidification or vitrification, and biological treatment.

estimate - a characteristic from the sample from which inferences on parameters can be made.

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. An alternative name is false negative decision error.

false negative decision error - see false acceptance decision error.

false positive decision error - see false rejection decision error.

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 decision error as alpha (α), the level of significance, or the size of the critical region, and it is expressed numerically as a probability. An alternative name is false positive decision error.

field variability - see sampling design error.

gray region - the range of possible parameter values near the action level where the cost of determining that the alterative 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."

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. Also known as measurement variability.

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.

null hypothesis - see baseline condition.

parameter - a description 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 100 ppm is the 25th percentile of a sample, then 25 percent of the dat are less than or equal to 10 ppm and 75 percent of the dat are greater than 10 ppm.

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 quality assurance and quality control 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.

precision - a measure of mutual agreement among individual measurements of the same property, usually under prescribed similar conditions expressed generally in terms of the standard deviation.

quality assurance (QA) - an integrated system of management activities involving planning, implementation, documentation, assessment, reporting, and quality improvement to ensure that a process, 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 quality assurance, quality control, 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 measure 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.

quality system - a structured and documented system describing the policies, objectives, principles, organizational authority, responsibilities, accountability, and implementation plan of an organization or ensuring quality in its work processes, products (items), and services. The quality system provides the framework for planning, implementing, documenting, and assessing work performed by the organization and for carrying out required quality assurance and quality control.

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

sample - (a) 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 (b) a group of samples from a statistical population whose properties are studies to gain information about the whole. The definition is decided by context of usage.

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 quality assurance and quality control 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. Also known as field variability.

stakeholder - a person or organization having an interest in the development of the project.

standard deviation - a measure of the dispersion or imprecision of a sample or population distribution expressed as the positive square root of the variance and has the same unit of measurement as the mean.

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

standard operating procedure - a written document that details the method for an operation, analysis, or action with thoroughly prescribed techniques and steps and that is officially approved as the method for performing certain routine or repetitive tasks.

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

total variability - see total study error.

type I error - the statistical term for false rejection decision error.

type II error - the statistical term for false acceptance decision error.

variability - refers to observed difference attributable to heterogeneity or diversity in a population. Sources of variability are the results of natural random processes and stem from environmental differences among the elements of the population. Variability is not usually reducible by further measurement but can be better estimated by increased sampling.

variance - a measure of a 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.