LANL ~~[illegible]~~ 94 9/1/94
Reference

GUIDANCE FOR THE DATA QUALITY OBJECTIVES PROCESS

EPA QA/G-4

United States Environmental Protection Agency
Quality Assurance Management Staff

Washington, DC 20460

FINAL

SEPTEMBER 1994

9628

FOREWORD

The U.S. Environmental Protection Agency (EPA) has developed the Data Quality Objectives (DQO) Process as an important tool for project managers and planners to determine the type, quantity, and quality of data needed to support Agency decisions. This guidance is the culmination of experiences in applying DQOs in different Program Offices at the EPA. Many elements of prior guidance, advice, statistics, and scientific planning have been incorporated into this document. This guidance supersedes all previous guidance, including the EPA's "Development of Data Quality Objectives, Description of Stages I and II" (July 1986), and "Guidance for Planning for Data Collection in Support of Environmental Decision Making Using the Data Quality Objectives Process" (Interim Final, October 1993). This document is consistent with the Office of Emergency and Remedial Response guidance, "Data Quality Objectives for Superfund" (EPA 540-R-93-071).

The purpose of this document is to provide general guidance to organizations on developing data quality criteria and performance specifications for decision making. This guidance assumes that an appropriate Quality System has been established and is operational.

This guidance has been prepared in response to EPA Order 5360.1, entitled "Policy and Program Requirements to Implement the Quality Assurance Program," which establishes requirements for quality assurance when generating environmental data in support of Agency decisions. In addition, this guidance reflects the policy of the Agency to develop and implement the DQO Process as expressed by Deputy Administrator A. James Barnes in his memorandum on "Agency Institutionalization of Data Quality Objectives," dated November 1986.

This document is a product of the collaborative effort of many quality management professionals throughout the EPA and among the contractor community. It has been peer reviewed by the EPA Program Offices, Regional Offices, and Laboratories. Many valuable comments and suggestions have been incorporated to make it more useful.

Table of Contents

Chapter		Page
	Foreword	ii
	List of Figures and Tables	iv
	Introduction	1
1.	Step 1: State the Problem	9
2.	Step 2: Identify the Decision	13
3.	Step 3: Identify the Inputs to the Decision	17
4.	Step 4: Define the Boundaries of the Study	19
5.	Step 5: Develop a Decision Rule	23
6.	Step 6: Specify Tolerable Limits on Decision Errors	27
7.	Step 7: Optimize the Design for Obtaining Data	37
	Bibliography	41

Appendices

A.	Beyond the DQO Process: The Quality Assurance Project Plan and Data Quality Assessment	43
B.	DQO Case Study: Cadmium-Contaminated Fly Ash Waste	47
C.	Derivation of Sample Size Formula for Testing Mean of Normal Distribution Versus an Action Level	61
D.	Glossary of Terms	65

List of Figures

Page

0-1. The Data Quality Objectives Process 2

0-2. Repeated Application of the DQO Process Throughout the Life Cycle of a Single Project 5

2-1. Example of Multiple Decisions Organized Into a Flowchart 16

4-1. An Example of How to Stratify a Site With Soil Contamination 22

6-1. An Example of a Decision Performance Goal Diagram — Baseline Condition: Parameter Exceeds Action Level 35

6-2. An Example of a Decision Performance Goal Diagram — Baseline Condition: Parameter is Less Than Action Level 36

7-1. An Example of a Power Curve — Baseline Condition: Parameter is Less Than Action Level 40

A-1. QA Planning and the Data Life Cycle 44

A-2. Quality Assurance Assessment 46

B-1. Design Performance Goal Diagram for Cadmium Compliance Testing — Baseline Condition: Mean Exceeds Action Level 53

List of Tables

1-1. Elements of the Problem Description 12

5-1. Attributes of Different Statistical Parameters to Characterize the Population 25

6-1. Decision Error Limits Table Corresponding to Figure 6-1 35

6-2. Decision Error Limits Table Corresponding to Figure 6-2 36

INTRODUCTION

Each year the U.S. Environmental Protection Agency (EPA) and the regulated community spend approximately $5 billion collecting environmental data for scientific research, regulatory decision making, and regulatory compliance. While these activities are necessary for effective environmental protection, it is the goal of EPA and the regulated community to minimize expenditures related to data collection by eliminating unnecessary, duplicative, or overly precise data. At the same time, the data collected should have sufficient quality and quantity to support defensible decision making. The most efficient way to accomplish both of these goals is to establish criteria for defensible decision making before the study begins, and then develop a data collection design based on these criteria. To facilitate this approach, the Quality Assurance Management Staff (QAMS) of EPA has developed the Data Quality Objectives (DQO) Process, a systematic planning tool based on the Scientific Method for establishing criteria for data quality and for developing data collection designs. By using the DQO Process to plan environmental data collection efforts, EPA can improve the effectiveness, efficiency, and defensibility of decisions in a resource-effective manner.

What are DQOs? DQOs are qualitative and quantitative statements derived from the outputs of the first six steps of the DQO Process that:

1) Clarify the study objective;

2) Define the most appropriate type of data to collect;

3) Determine the most appropriate conditions from which to collect the data; and

4) Specify tolerable limits on decision errors which will be used as the basis for establishing the quantity and quality of data needed to support the decision.

The DQOs are then used to develop a scientific and resource-effective data collection design.

What is the DQO Process? The DQO Process is a strategic planning approach based on the Scientific Method that is used to prepare for a data collection activity. It provides a systematic procedure for defining the criteria that a data collection design should satisfy, including when to collect samples, where to collect samples, the tolerable level of decision errors for the study, and how many samples to collect.

By using the DQO Process, the Agency will assure that the type, quantity, and quality of environmental data used in decision making will be appropriate for the intended application. In addition, the Agency will guard against committing resources to data collection efforts that do not support a defensible decision.

The DQO Process consists of seven steps, as shown in Figure 0-1. The output from each step influences the choices that will be made later in the Process. Even though the DQO Process is depicted as a linear sequence of steps, in practice it is iterative; the outputs from one step may lead to reconsideration of prior steps. This iteration should be encouraged since it will ultimately lead to a more efficient data collection design. During the first six steps of the DQO Process, the planning team will develop the decision performance criteria (DQOs) that will be used to develop the data collection design. The final step of the Process involves developing the data collection design based on the DQOs. The first six steps should be completed before the planning team attempts to develop the data collection design because this final step is dependent on a clear understanding of the first six steps taken as a whole. In Figure 0-1, the iterative link between the DQOs and the Optimize the Design step is illustrated by double arrows, which signify that it may be necessary to revisit any one or more of the first six steps to develop a feasible and appropriate data collection design. Above all, every step should be completed before data collection begins.

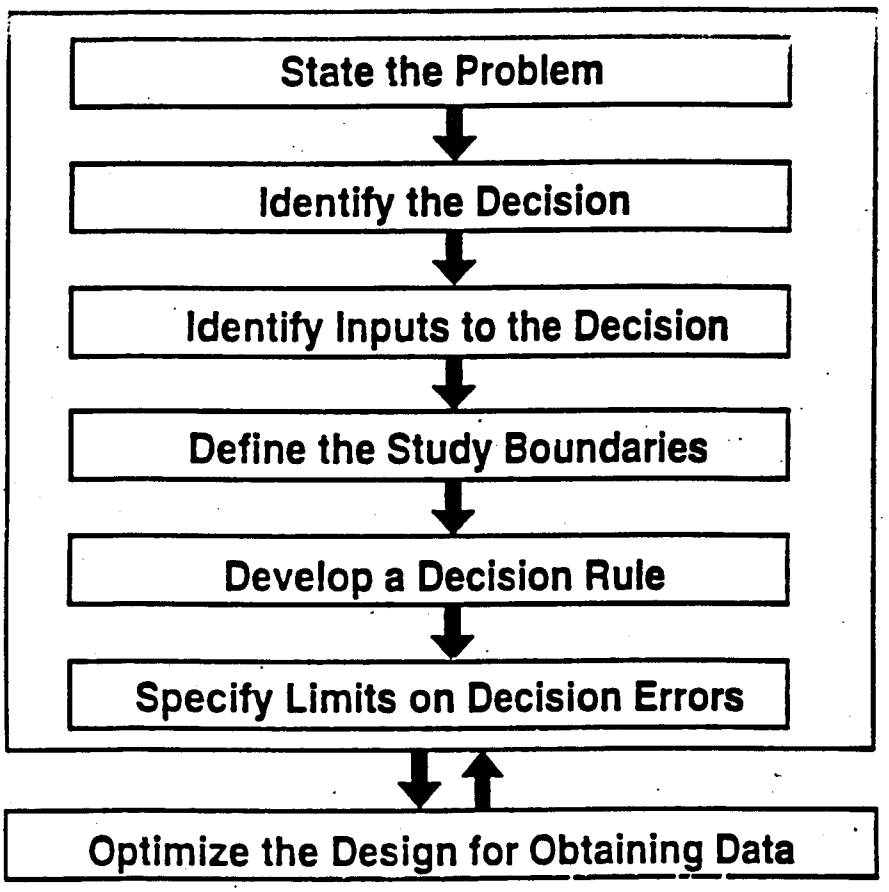


Figure 0-1. The Data Quality Objectives Process.

Each of the seven steps is described briefly below. A more detailed description can be found in the subsequent chapters of this guidance.

- Step 1: State the Problem — Concisely describe the problem to be studied. Review prior studies and existing information to gain a sufficient understanding to define the problem.

- Step 2: Identify the Decision — Identify what questions the study will attempt to resolve, and what actions may result.

- Step 3: Identify the Inputs to the Decision — Identify the information that needs to be obtained and the measurements that need to be taken to resolve the decision statement.

- Step 4: Define the Study Boundaries — Specify the time periods and spatial area to which decisions will apply. Determine when and where data should be collected.

- Step 5: Develop a Decision Rule — Define the statistical parameter of interest, specify the action level, and integrate the previous DQO outputs into a single statement that describes the logical basis for choosing among alternative actions.

- Step 6: Specify Tolerable Limits on Decision Errors — Define the decision maker's tolerable decision error rates[1] based on a consideration of the consequences of making an incorrect decision.

- Step 7: Optimize the Design — Evaluate information from the previous steps and generate alternative data collection designs. Choose the most resource-effective design that meets all DQOs.

Who should read the DQO guidance? This guidance is intended for project managers and other members of a planning team that will use the DQO Process to structure the data collection planning process and to develop an appropriate data collection design. In addition, the guidance may be relevant to other staff members who will participate in the study. Consult with an EPA Quality Assurance Manager, Quality Assurance Officer, or Quality Assurance Representative to obtain additional advice on who should read this guidance.

[1] A decision error rate is the probability of making an incorrect decision based on data that inaccurately estimate the true state of nature.

What projects are covered by this guidance? This guidance document covers all projects where:

1) the objective of the study is to collect environmental data in support of an Agency program, and

2) the results of the study will be used to make a specific decision.

Every step of this guidance may not be applicable to data collection activities where specific decisions cannot be identified, such as studies that are exploratory in nature. The reason for this distinction is that part of the DQO Process includes formulating statistical hypotheses. If a statistical hypothesis is not linked to a clear decision in which the decision maker can identify potential consequences of making a decision error, then some of the activities recommended in this guidance may not apply. Nonetheless, the DQO Process is still a valuable tool that can be used to help plan studies where the data are not directly used to support a specific decision. In these cases, it may be possible to frame a research type study question in the form of a decision or modify the activities described in this guidance to address the needs of the study.

What is the value of using the DQO Process?

- 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 structure of the DQO Process provides a convenient way to document activities and decisions and to communicate the data collection design to others.

- The DQO Process enables data users and relevant technical experts to participate 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.

- 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 probability 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 controlling decision errors and satisfying the DQOs.

- 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.

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. In general, EPA's policy is to use the DQO Process to plan all data collection efforts that will require or result in a substantial commitment of resources. The Quality Management Plans (QMPs) of the Agency's National Program Offices, Regional Offices, and Research and Development organizations will specify which studies require DQOs.

Can the DQO Process be used for small studies? The DQO Process applies to any study, regardless of its size. However, the depth and detail of DQO development will depend on the complexity of the study. The more complex a study, the more likely that it will have several decisions that could benefit from the DQO Process and that the decisions will require more intensive DQO development.

Should the DQO Process be applied as intensively to all situations? No, the DQO Process is a flexible planning tool that can be used more or less intensively as the situation requires. For projects that have multiple decisions, where the resolution of one decision only leads to the evaluation of subsequent decisions, 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 possibility of making a decision error becomes more critical, however, the level of effort needed to resolve a decision generally will become greater. Figure 0-2 illustrates this point.

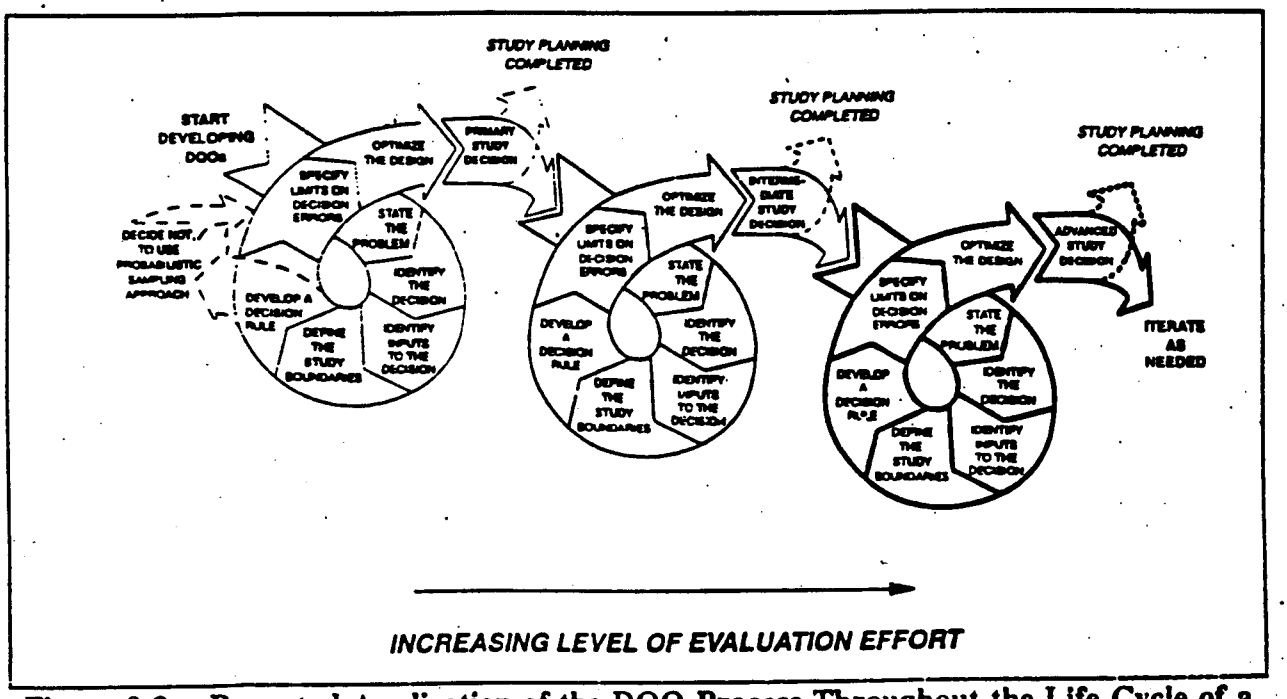


Figure 0-2. Repeated Application of the DQO Process Throughout the Life Cycle of a Single Project.

Who participates in the DQO Process? A DQO planning team generally consists of senior program staff, technical experts, senior managers, someone with statistical expertise, and a Quality Assurance (QA)/Quality Control (QC) advisor, such as a QA Manager. It is important that all of these people, including managers, participate (or stay informed) from the beginning of the DQO Process so that it can proceed efficiently.

What are the outputs of the DQO Process? The DQO Process leads to the development of a quantitative and qualitative framework for a study. Each step of the Process derives valuable criteria that will be used to establish the final data collection design. The first five steps of the DQO Process identify mostly qualitative criteria such as what problem has initiated the study and what decision it attempts to resolve. They also define the type of data that will be collected, where and when the data will be collected, and a decision rule that defines how the decision will be made. The sixth step defines quantitative criteria expressed as limits on decision errors that the decision maker can tolerate. The final step is used to develop a data collection design based on the criteria developed in the first six steps. The final product of the DQO Process is a data collection design that meets the quantitative and qualitative needs of the study.

Much of the information that is developed in the DQO Process will also be useful for the development of Quality Assurance Project Plans (QAPPs) and the implementation of the Data Quality Assessment (DQA) Process. The outputs of the DQO Process can be used directly and indirectly as inputs to a QAPP. To evaluate the data using the DQA Process, it is necessary to have first established decision quality criteria using the DQO Process or its equivalent. Therefore, the DQO Process not only helps plan a study, establish decision quality criteria, and develop a data collection design, but it also aids in the development of QAPPs and the DQA Process.

What is a data collection design? A data collection design specifies the final configuration of the environmental monitoring or measurement effort required to satisfy the DQOs. It designates the types and quantities of samples or monitoring information to be collected; where, when, and under what conditions they should be collected; what variables are to be measured; and the QA/QC procedures to ensure that sampling design and measurement errors are controlled sufficiently to meet the tolerable decision error rates specified in the DQOs. These QA/QC procedures are established in the QAPP.

Where does the DQO Process fit into EPA's Quality System? The DQO Process is the part of the Quality System that provides the basis for linking the intended use of the data to the QA/QC requirements for data collection and analysis. This document is one of a series of quality management requirements and guidance documents that the U.S. EPA Quality Assurance Management Staff (QAMS) has prepared to assist users in implementing the Agency-wide Quality System. The current document list contains:

EPA QA/R-1 EPA Quality System Requirements for Environmental Programs

EPA QA/G-1 Guidance for Developing, Implementing, and Evaluating Quality Systems for Environmental Programs

EPA QA/R-2 EPA Requirements for Quality Management Plans

EPA QA/G-2 Guidance for Preparing Quality Management Plans for Environmental Programs

EPA QA/G-4 Guidance for The Data Quality Objectives Process

EPA QA/R-5 EPA Requirements for Quality Assurance Project Plans for Environmental Data Operations

EPA QA/G-5 Guidance for Quality Assurance Project Plans

EPA QA/G-9 Guidance for Data Quality Assessments

Agency policy statements are found in the requirements documents (QA/R-xx series). Advisory papers are found in the guidance documents (QA/G-xx series).

Can existing data be used to support decisions using the DQO Process? Existing data can be very useful for supporting decisions using the DQO Process. There are three ways that existing data can be used:

1) If sufficient documentation is available, existing data may be used alone or combined with new data. Determining whether data can appropriately be combined can be a very complex operation that should be undertaken with great care. In many cases it will require the expertise of a statistician.

2) The existing data may provide valuable information (such as variability) that can be used in the development of the data collection design.

3) The existing data may be useful in guiding the selection of an efficient data collection design.

Will the use of the DQO Process always result in statistical/probabilistic sampling methods for data collection? No. While statistical methods for developing the data collection design are strongly encouraged, this guidance recognizes that not every problem can be evaluated using probabilistic techniques. The DQO Process, however, 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 on how to evaluate the result of the data collection. When non-probabilistic, judgemental, or quota sampling methods are used, be sure to consult with an EPA QA Manager, QA Officer, or QA Representative to ensure that program-specific QA requirements are satisfied.

How should this guidance be used? This guidance should be used 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 technical experts, program managers, and decision makers.

How is this guidance structured? This guidance contains seven chapters, four appendices, and a bibliography. Each of the remaining chapters describes one of the seven steps of the DQO Process. Each chapter is divided into four sections as follows:

(1) Purpose - This section explains the objective of the chapter.

(2) Expected Outputs - This section identifies the products expected upon completion of the DQO Process step.

(3) Background - This section provides background information on the DQO Process step, including the rationale for the activities in that step.

(4) Activities - This section describes the activities recommended for completing the DQO Process step, including how inputs to the step are used.

Appendix A provides a brief overview of both the Quality Assurance Project Plan (QAPP) development process, which is used to document the operational and QA/QC procedures needed to implement the data collection design, and the Data Quality Assessment (DQA) Process, which is used after the data have been collected to evaluate whether the DQOs have been satisfied. Appendix B is a case study in which the DQO Process is applied to an environmental problem. Appendix C provides a derivation of the sample size formula used in Appendix B. Appendix D provides a glossary of terms used in this guidance.

Where is it possible to get statistical support? Access to statistical support is available through the EPA Quality Assurance Management Staff (QAMS) at (202) 260-5763.

How long will this guidance be in effect? This guidance will remain in effect for five years from the publication date, unless superseded by an updated version.

Where is it possible to get more information about the DQO Process? A DQO training course is available through the EPA at the U.S. EPA Headquarters in Washington, D.C. Additional documents on DQO applications can be obtained from the Quality Assurance Management Staff at EPA Headquarters.

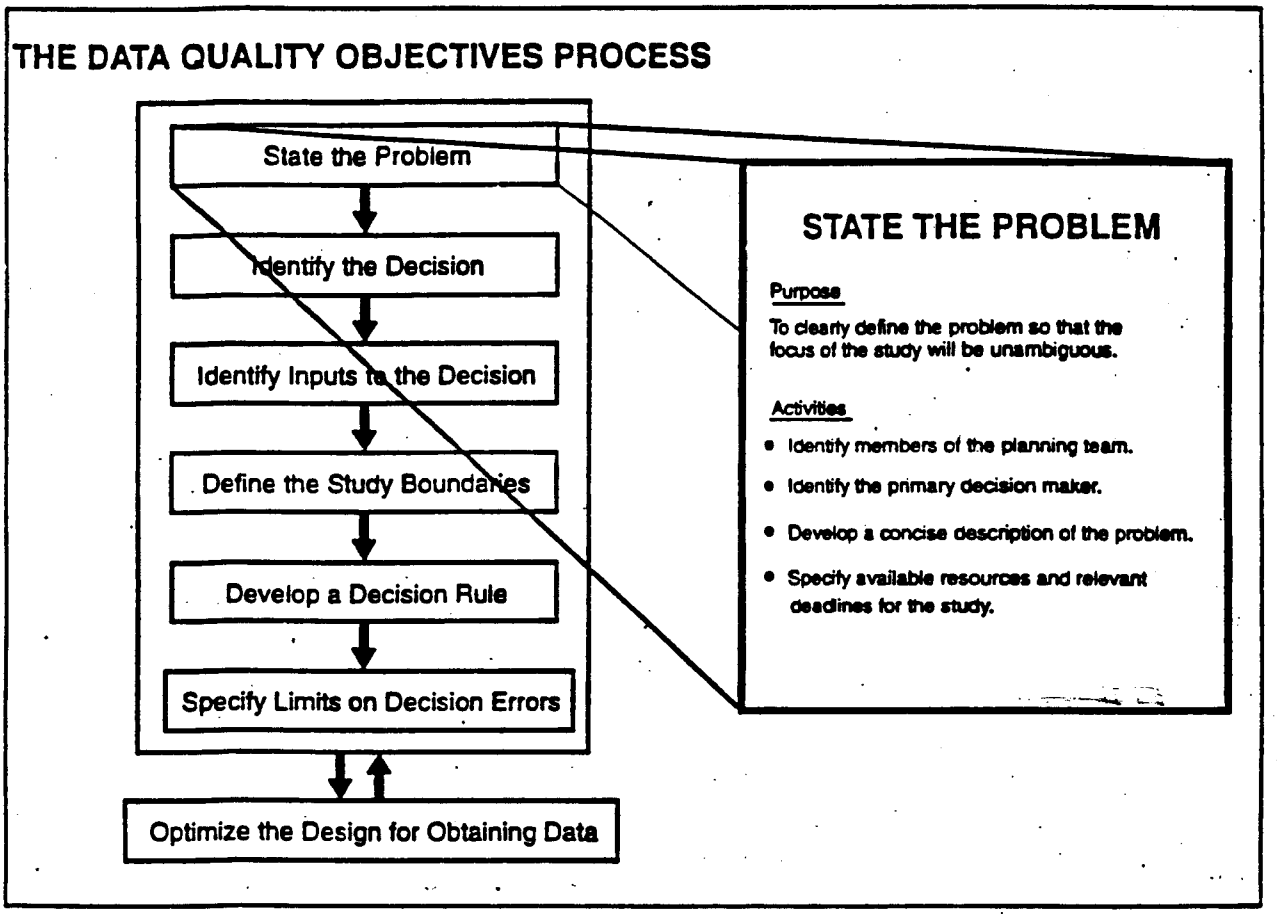
Two documents that can provide additional detail on the DQO Process are:

- U.S. Environmental Protection Agency. 1993. *Data Quality Objectives Process for Superfund: Interim Final Guidance*. EPA 540-R-93-071.

- Bates, D.J., R.O. Gilbert, N.L. Hassig, R.F. O'Brien, B.A. Pulsipher, 1993. Decision Performance Criteria: The Driver Behind The Data Quality Objectives Process — A Statistical Introduction (Draft). Pacific Northwest Laboratory, Richland, Washington.

CHAPTER 1

STEP 1: STATE THE PROBLEM



Purpose

The purpose of this step is to define the problem so that the focus of the study will be unambiguous.

Expected Outputs

- A list of the planning team members and identification of the decision maker.
- A concise description of the problem.
- A summary of available resources and relevant deadlines for the study.

Background

The first step in any decision making process is to define the problem that has initiated the study. Since most environmental problems present a complex interaction of technical, economic, social, and political factors, it is critical to the success of the process to define the problem completely and in an uncomplicated format. A problem will have the greatest chance of being solved when a multidisciplinary team of technical experts and stakeholders can help to recognize all of the important facets of the problem and ensure that complex issues are described accurately. Generally teams will function more effectively when they have one clearly identified decision maker.

This step in the DQO Process addresses development of a planning team that will define the problem and implement subsequent steps of the Process. It also calls for the identification of a decision maker who will lead the planning team and make final resolutions during the Process. The goal is to create a well-structured planning team that will work effectively and efficiently to develop a concise and complete description of the problem, which will provide the basis for the rest of the DQO development.

Activities

Identify members of the planning team. The planning team is the group that will develop DQOs for the study. The number of planning team members will be directly related to the size and complexity of the problem. The team should include representatives from all groups who are stakeholders in the project, including, but not limited to, samplers, chemists and other scientists and engineers, modelers, technical project managers, community representatives, administrative and executive managers, QA/QC experts (such as a QA Manager), data users, and decision makers. A reasonable effort should be made to include any decision makers who may use the study findings later. A statistician (or someone knowledgeable and experienced with environmental statistical design and analysis) should also be included on this team.

Identify the primary decision maker of the planning team and define each member's role and responsibility during the DQO Process. The planning team generally has a leader, referred to as the "decision maker." The decision maker has the ultimate authority for making final decisions based on the recommendations of the planning team. The decision maker is often the person with the most authority over the study, and may be responsible for assigning the roles and responsibilities to the planning team members. In cases where the decision maker cannot attend DQO planning meetings, a senior staff member should keep the decision maker informed of important planning issues.

Develop a concise description of the problem. The problem description provides background information on the fundamental issue to be addressed by the study. Below is a list of steps that may be helpful during this phase of DQO development.

- Describe the conditions or circumstances that are causing the problem and the reason for understanding the study. Typical examples for environmental problems include conditions that may pose a threat to human health or the environment, and circumstances of potential non-compliance with regulations.

- Describe the problem as it is currently understood by briefly summarizing existing information. (See Table 1-1 for a list of elements that may be appropriate to include in the problem description.)

- Conduct literature searches and examine past or ongoing studies to ensure that the problem is correctly defined and has not been solved previously. Organize and review relevant information, including preliminary studies, and indicate the source and reliability of the information. Take note of information about the performance of sampling and analytical methods observed in similar studies since this information may prove to be particularly valuable later in the DQO Process.

- If the problem is complex, consider breaking it into more manageable pieces. Identify those pieces that could be addressed by separate studies. Assign priorities to and logical relationships among the pieces of the problem.

Specify the available resources and relevant deadlines for the study. Stipulate the anticipated budget, available personnel, and contractual vehicles (if applicable). Also, enumerate any deadlines for completion of the study and any intermediate deadlines that may need to be met.

Elements of the Problem Description

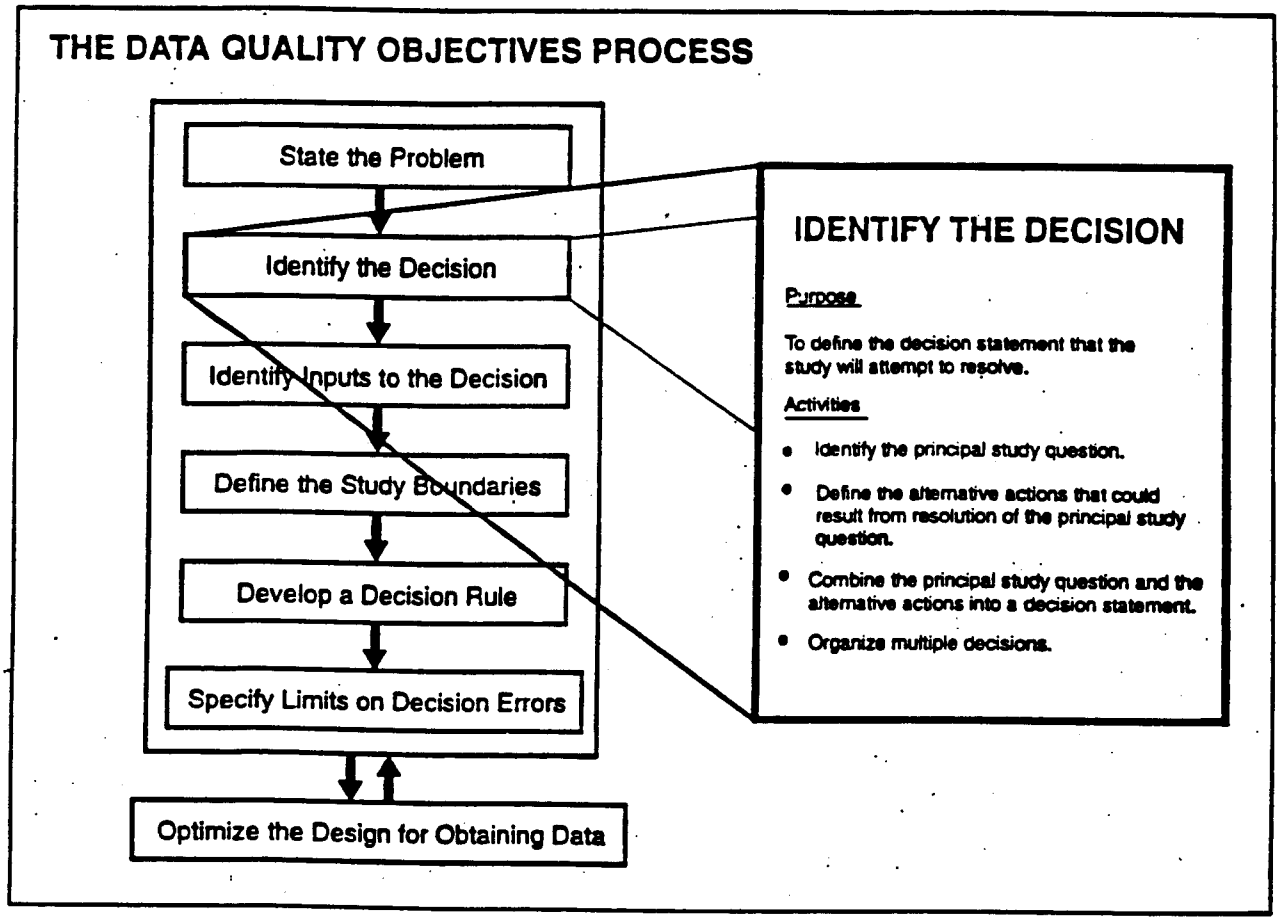
The following elements may be appropriate to include in the problem description. Note: this list only provides the basic elements of the problem description. Your elements may be slightly different.

- Study objectives/regulatory context.
- Persons or organizations involved in the study.
- Persons or organizations that have an interest in the study.
- Political issues surrounding the study.
- Sources and amount of funding.
- Previous study results.
- Existing sampling design constraints (some aspects of sampling design may be specified in regulations or established through past planning efforts).

Table 1-1. Elements of the Problem Description.

CHAPTER 2

STEP 2: IDENTIFY THE DECISION



Purpose

The purpose of this step is to define the decision statement that the study will attempt to resolve.

Expected Outputs

- **A decision statement that links the principal study question to possible actions that will solve the problem.**

Background

The goal of this step is to define the question that the study will attempt to resolve and identify the alternative actions that may be taken based on the outcome of the study. In the DQO Process the combination of these two elements is called the decision statement or decision. The decision statement is critical for defining decision performance criteria later in the Process.

The three activities in this chapter usually are most easily developed in the order that they appear. Sometimes, however, it is easier to identify alternative actions before the principal study question. In these cases, identify alternative actions that address the problem, then define the principal study question.

In some cases, several decision statements are appropriate to address the problem under investigation. In these instances, the planning team should organize the decision statements in order of priority and identify the most logical and efficient sequence for analyzing and resolving them. If the principal study question is not obvious and specific alterative actions cannot be identified, then the study may fall in the category of exploratory research, in which case this step of the DQO Process may not be applicable.

Activities

Identify the principal study question. Based on a review of the problem stated in Step 1, identify the principal study question and state it as specifically as possible. A specific statement of the principal study question narrows 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. The following examples illustrate this point:

- "Is the permittee out of compliance with discharge limits?"
- "Does the pollutant concentration exceed the National Ambient Air Quality Standard?"
- "Is the contaminant concentration significantly above background levels (which would indicate that a release has occurred)?"

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

Define the alternative actions that could result from resolution of the principal study question. Identify the possible actions that may be taken to solve the problem, including the alternative that does not require action. The types of actions considered will depend logically on the possible answers to the principal study question. These alternative actions form the basis for defining decision performance criteria in Step 6: Specify Tolerable Limits on Decision Errors.

The following example illustrates how alternative actions are defined based on possible answers to the following principal study question: "Are the lead pellets that are fired by bird hunters and collect on the bottom of ponds contributing to the decrease in the duck population in Adelayed County?" Possible resolutions of the principal study question are 1) the lead pellets are a factor in the decrease of the duck population, or 2) the lead pellets are not a factor in the duck population's decrease. If the lead is a contributing factor, the action may be to remove the lead from the bottom of the ponds and, at the same time, regulate the type of pellets that hunters may use in the future. If lead pellets are not found to contribute to a decrease in the duck population, then no action will be taken.

Combine the principal study question and the alternative actions into a decision statement. Combine the alternative actions identified in the previous activity and the principal study question into a decision statement that expresses a choice among alternative actions. The following standard form 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]."

To illustrate the decision statement framing activity, consider the previous example. The principal study question is, "Are lead pellets on the bottom of ponds in Adelayed County contributing to the decrease in the duck population?", and the alternative actions are to "remediate the lead and regulate the use of lead pellets for hunting," or "take no action." Therefore the decision statement is, "Determine whether or not lead pellets are contributing to the decrease in the duck population and require remediation and regulation." For a compliance monitoring problem, a decision statement that incorporates the principal study question and expresses a choice among alternative actions might be, "Determine whether or not the permittee is out of compliance with discharge limits and requires enforcement action."

Organize multiple decisions. If several separate decision statements must be defined to address the problem, list them and identify the sequence in which they should be resolved. It may be useful to document the decision resolution sequence and relationships in a diagram or flowchart (see example in Figure 2-1).

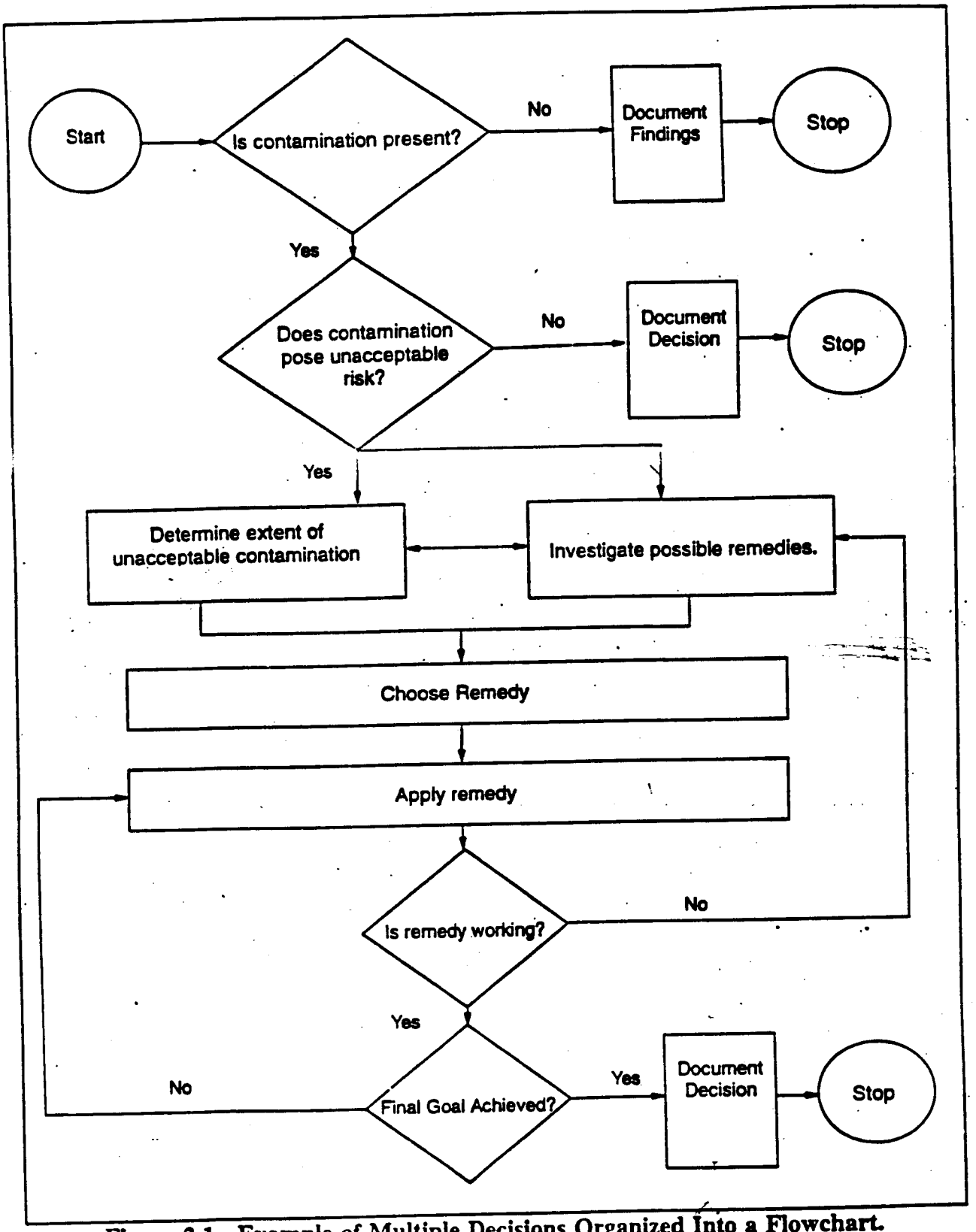
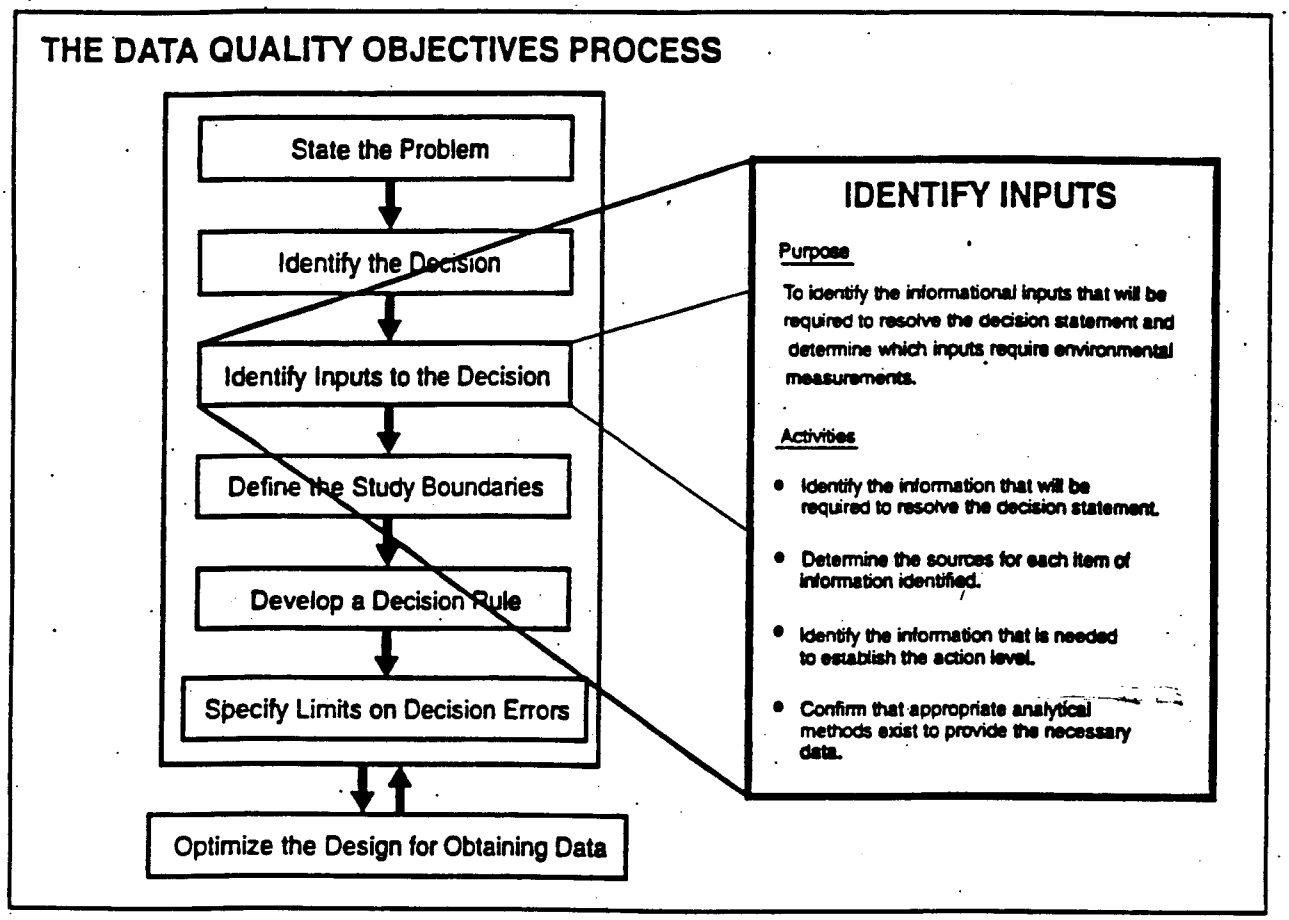


Figure 2-1. Example of Multiple Decisions Organized Into a Flowchart.

CHAPTER 3

STEP 3: IDENTIFY THE INPUTS TO THE DECISION



Purpose

The purpose of this step is to identify the informational inputs that will be required to resolve the decision statement and determine which inputs require environmental measurements.

Expected Outputs

- **A list of informational inputs needed to resolve the decision statement.**
- **A list of environmental variables or characteristics that will be measured.**

Background

To resolve most decision statements, it is necessary to collect data or information. In this step, the planning team identifies the different types of information that will be needed to resolve the decision statement. The key information requirements include the measurements that may be required, the source of data or information (e.g., historic or new data), and the basis for setting the action level. Once the planning team has determined what needs to be measured, they will refine the specifications and criteria for these measurements in later steps of the DQO Process.

Activities

Identify the information that will be required to resolve the decision statement. Determine which environmental variables or other information are needed to resolve the decision statement. Consider whether monitoring or modeling approaches, or a combination of both, will be used to acquire the information. Based on the selected data acquisition approach, identify the types of information needed to support the decision statement. Ask general questions such as, "Is information on the physical properties of the media required?" or "Is information on the chemical characteristics of the matrix needed?" These types of questions and their answers help identify the information needs. In compliance monitoring for pollutants discharged into surface water, examples of environmental variables of interest may include levels of lead, silver, total suspended solids, or temperature measurements.

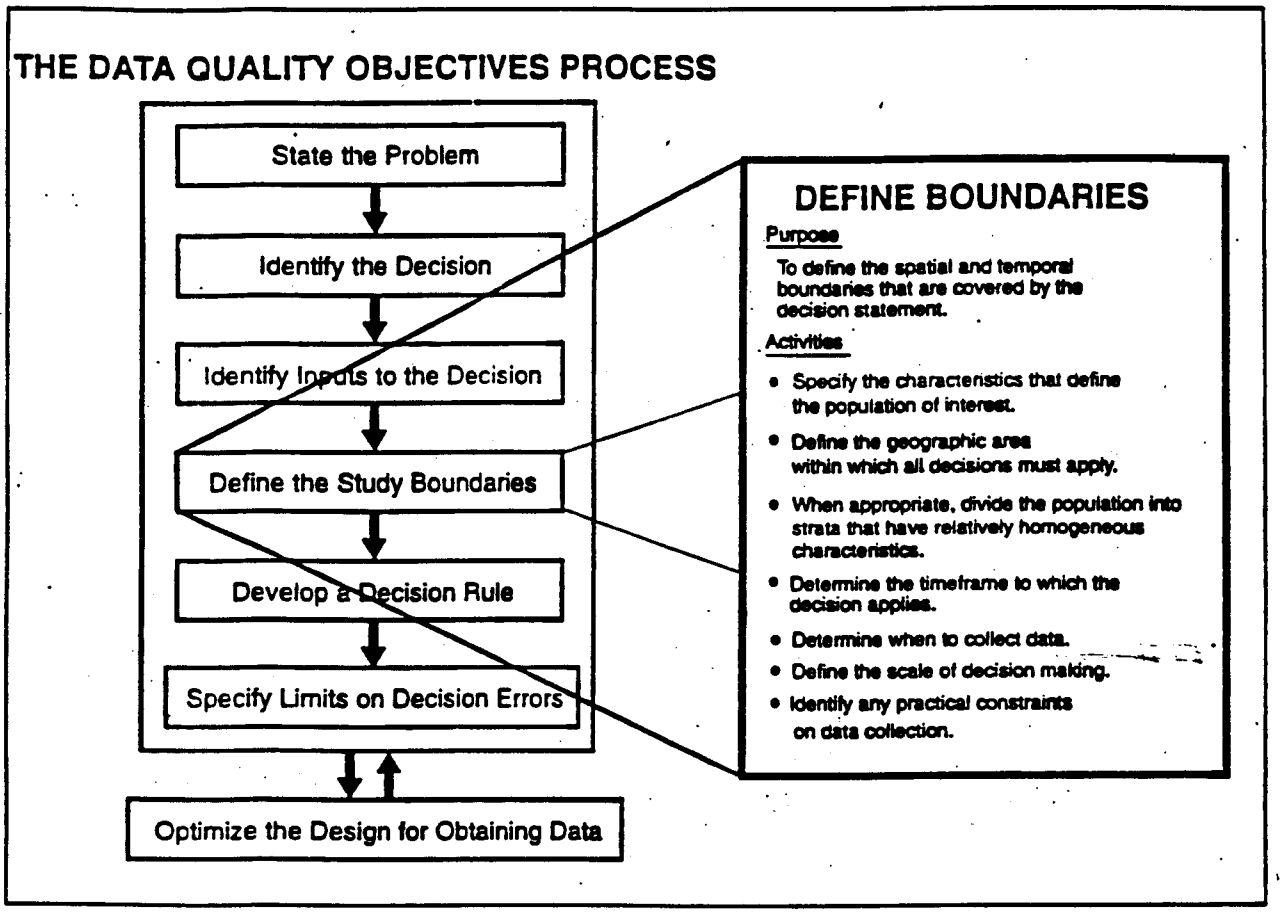
Determine the sources for each item of information identified above. Identify and list the sources for the information needed to resolve the decision statement. These sources may include results of previous data collections, historical records, regulatory guidance, professional judgement, scientific literature, or new data collections. Next, qualitatively evaluate whether any existing data are appropriate for the study. Existing data will be evaluated quantitatively in Step 7: Optimize the Design for Obtaining Data.

Identify the information that is needed to establish the action level. Define the basis for setting the action level. The action level is the threshold value which provides the criterion for choosing between alternative actions. Action levels may be based on regulatory thresholds or standards, or they may be derived from problem-specific considerations such as risk analysis. In this step, simply determine the criteria that will be used to set the numerical value. The actual numerical action level will be set in Step 5: Develop a Decision Rule.

Confirm that appropriate measurement methods exist to provide the necessary data. Use the list of environmental measurements identified earlier in this step to develop a list of potentially appropriate measurement methods. Note the method detection limit and limit of quantitation for each potential method; this performance information will be used in steps 5 and 7 of the DQO Process.

CHAPTER 4

STEP 4: DEFINE THE BOUNDARIES OF THE STUDY



Purpose

The purpose of this step is to define the spatial and temporal boundaries of the problem.

Expected Outputs

- **A detailed description of the spatial and temporal boundaries of the problem.**
- **Any practical constraints that may interfere with the study.**

Background

It is difficult to interpret data that have not been drawn from a well-defined population. The term "population" refers to the total collection or universe of objects or people to be studied, from which samples will be drawn. The purpose of this step is to define spatial and temporal components of the population that will be covered by the decision statement so that the data can be easily interpreted. These components include:

- Spatial boundaries that define the physical area to be studied and from where the samples should be taken, and
- Temporal boundaries that describe the timeframe the study data will represent and when the samples should be taken.

The boundaries will be used to ensure that the data collection design incorporates the time periods in which the study should be implemented, areas that should be sampled, and the time period to which the study results should apply. This will help ensure that the study data are representative of the population being studied. Defining boundaries before the data are collected can also prevent inappropriate pooling of data in a way that masks useful information.

Practical constraints that could interfere with sampling should also be identified in this step. A practical constraint is any hinderance or obstacle that potentially may interfere with the full implementation of the data collection design.

Activities

Specify the characteristics that define the population of interest. Specify the characteristics that define the population. It is important to clearly define the attributes that make up the population by stating them in a way that makes the focus of the study unambiguous. For example, the population may be PCB concentrations in soil, lead concentrations in the blood of children under the age of seven, or hourly ozone concentrations within the metropolitan area. There may be several ways to define a population; always choose the one that is most specific. For example, "tetrachlorodibenzodioxin" is more specific than "dioxin," and "hexavalent chromium" is more specific than "chromium".

Define the spatial boundary of the decision statement.

> **Define the geographic area to which the decision statement applies.** The geographic area is a region distinctively marked by some physical features (i.e., volume, length, width, boundary). Some examples of geographic areas are the metropolitan city limits, the soil within the property boundaries down to a depth of six inches, or the natural habitat range of a particular animal species.

When appropriate, divide the population into strata that have relatively homogeneous characteristics. Using existing information, stratify or segregate the elements of the population into subsets or categories that exhibit relatively homogeneous properties or characteristics that may have an influence on the outcome of the study, such as contaminant concentrations, age, or height. Dividing the population into strata is desirable for studying sub-populations, reducing variability within subsets of data, or reducing the complexity of the problem by breaking it into more manageable pieces. See Figure 4-1 for an example of how to stratify a site with soil contamination.

Define the temporal boundary of the problem.

Determine the timeframe to which the decision applies. It may not be possible to collect data over the full time period to which the decision will apply. Therefore the planning team should determine the timeframe that the data should reflect; for example, "The data will reflect the condition of contaminant leaching into ground water over a period of a hundred years," or "The data will be used to reflect the risk conditions of an average resident over their average length of residence which is estimated to be eight years." Timeframes should be defined for the overall population and any sub-populations of interest.

Determine when to collect data. Conditions may vary over the course of a study, which may affect the success of data collection and the interpretation of data results. These factors may include weather, temperature, humidity, or amount of sunlight and wind. Determine when conditions will be most favorable for collecting data and select the most appropriate time period to collect data that reflect those conditions. For example, a study to measure ambient airborne particulate matter may give misleading information if the sampling is conducted in the wetter winter months rather than the drier summer months.

Define the scale of decision making. Define the smallest, most appropriate subsets of the population (sub-populations) for which decisions will be made based on the spatial or temporal boundaries. 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 six inches of soil within the property boundaries, and the *population* is the lead concentration in surface soil. The *scale of decision making* could be set to an area which has a 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 site at this scale will be protective of children, a sensitive population in risk assessment. A temporal scale of decision making might be necessary for other types of studies. For example, in order to regulate water quality, it would be useful to set a scale of decision making that limits the time between sampling events. This would minimize the potential adverse effects in case the water quality was degraded between sampling events.

Identify any practical constraints on data collection. Identify any constraints or obstacles that could potentially interfere with the full implementation of the data collection design, such as seasonal or meteorological conditions when sampling is not possible, the inability to gain site access or informed consent, or the unavailability of personnel, time, or equipment. For example, it may not be possible to take surface soil samples beyond the east boundaries of a site under investigation because permission had not been granted by the owner of the adjacent property.

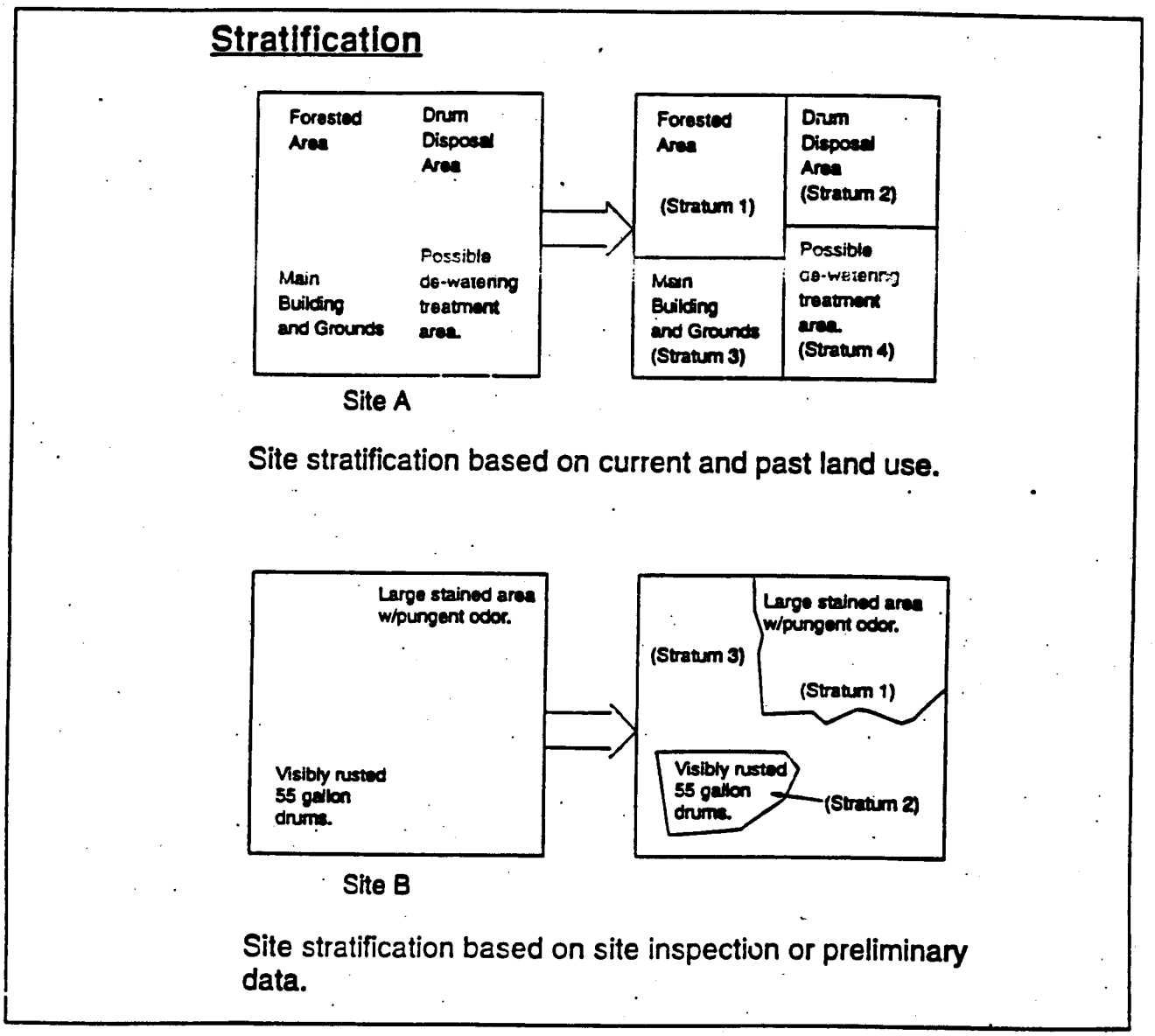
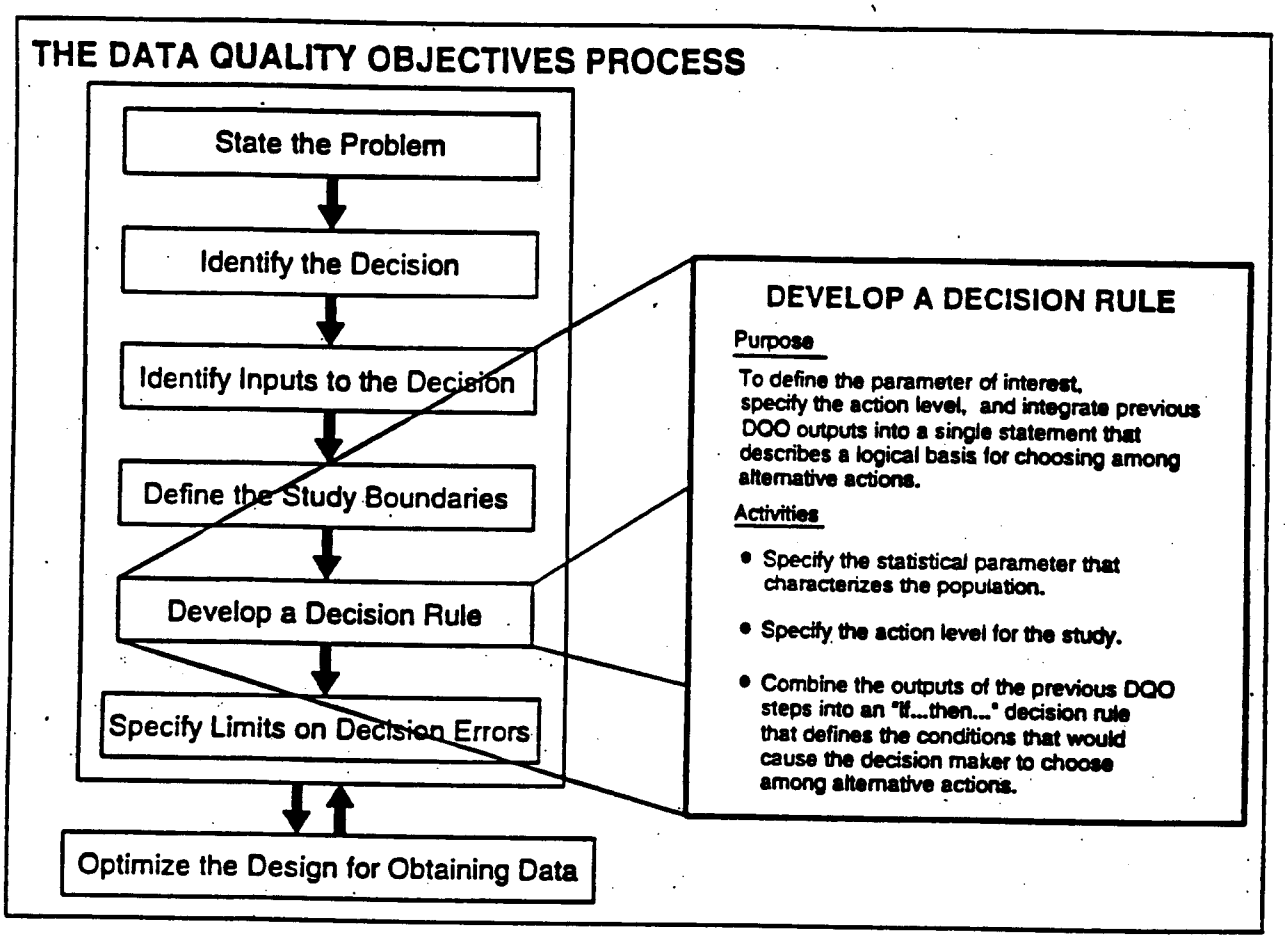


Figure 4-1. An Example of How to Stratify a Site with Soil Contamination.

CHAPTER 5

STEP 5: DEVELOP A DECISION RULE



Purpose

The purpose of this step is to define the parameter of interest, specify the action level, and integrate previous DQO outputs into a single statement that describes a logical basis for choosing among alternative actions.

Expected Outputs

- **The statistical parameter (the parameter of interest) that characterizes the population.**
- **The action level.**
- **An "if...then..." statement that defines the conditions that would cause the decision maker to choose among alternative actions.**

Background

The decision rule summarizes what attributes the decision maker wants to know about the population and how that knowledge would guide the selection of a course of action to solve the problem. The Decision Rule step combines criteria from past steps with the *parameter of interest* (statistical characteristic of the population) and the *action level* to provide a concise description of what action will be taken based on the results of the data collection.

There are four main elements to a decision rule:

(1) The *parameter of interest*, a descriptive measure (such as a mean, median, or proportion) that specifies the characteristic or attribute that the decision maker would like to know about the statistical population. The purpose of the data collection design is to produce environmental data that can be used to develop a reasonable estimate of the population parameter.

(2) The *scale of decision making*, the smallest, most appropriate subset (sub-population) for which separate decisions will be made. (The scale of decision making was defined in Step 4: Define the Boundaries of the Study.)

(3) The *action level*, a measurement threshold value of the parameter of interest that provides the criterion for choosing among alternative actions. The action level can be based on regulatory standards, an exposure assessment, technology based limits, or reference-based standards.

(4) The *alternative actions*, the actions that the decision maker would take, depending on the true value of the parameter of interest. (The alternative actions were identified in Step 2: Identify the Decision.)

Activities

Specify the statistical parameter that characterizes the population (the parameter of interest). The planning team should specify the parameter of interest (such as the mean, median, or percentile) whose true value the decision maker would like know and that the data will estimate. For example, to determine if the contamination level at a given site exceeds an action level, the planning team must specify the parameter that will be evaluated with respect to the action level (e.g., the mean concentration). Some regulations specify the parameter, but if this is not the case, it may be necessary to consult with a statistician to help select a parameter that is consistent with the intended application. Recognize that the parameter that is chosen in this step may be changed to an equivalent descriptive measure as more information becomes available based on statistical considerations in Step 7 of the DQO Process and in the Data Quality Assessment Process. Information about positive and negative attributes of commonly used parameters is provided at the end of this chapter.

Specify the action level for the study. The decision maker should specify the numerical value that would cause him/her to choose between alternative actions. For example, the decision maker would choose one action if the true value of the parameter of interest is above 1 mg/L, and a different action otherwise. Confirm that the action level is greater than the detection and quantitation limits for the potential measurement methods identified in Step 3: Identify the Inputs to the Decision.

Develop a decision rule. Develop a decision rule as an "if...then..." statement that incorporates the parameter of interest, the scale of decision making, the action level, and the action(s) that would result from resolution of the decision. These four elements are combined in the following way: If the parameter of interest (e.g., true mean concentration of lead) within the scale of decision making (e.g., 1-acre plots) is greater than the action level (e.g., 1 mg/Kg), then take alternative action A (e.g., remove the soil from the site); otherwise take alternative action B (e.g., leave the soil in place). For example, "If the true mean concentration of cadmium in the fly ash leachate within a container truck exceeds 1.0 mg/Kg, then the waste ash will be considered hazardous and will be disposed of in a RCRA hazardous waste landfill; otherwise, the waste ash will be disposed of in a municipal landfill." This statement is a functional decision rule that expresses what the decision maker ideally would like to resolve. It is not an operational decision rule which incorporates the decision maker's tolerable limits on decision errors and the statistical hypothesis, and describes how the data will be summarized. The operational decision rule is developed during the Data Quality Assessment Process, after the data have been collected (see Appendix A).

Attributes of Different Statistical Parameters

MEAN

Positive Attributes

- Useful when action level is based on long-term, average health effects (chronic conditions, carcinogenicity).
- Useful when the population is uniform with relatively small spread.
- Generally requires fewer samples than other parameters.

Negative Attributes

- Not a very representative measure of central tendency for highly skewed populations.
- Not useful when the population contains a large proportion of values that are less than measurement detection limits.

(continued)

Table 5-1. Attributes of Different Statistical Parameters to Characterize the Population

Attributes of Different Statistical Parameters (continued)

MEDIAN

Positive Attributes

- Useful when action level is based on long-term, average health effects (chronic conditions, carcinogenicity).
- Provides a more representative measure of central tendency than the mean for skewed populations.
- Useful when the population contains a large number of values that are less than measurement detection limits.
- Relies on few statistical assumptions.

Negative Attributes

- Will not protect against the effect of extreme values.
- Not a very representative measure of central tendency for highly skewed populations.

UPPER PROPORTION/PERCENTILE

Positive Attributes

- Useful for protection against extreme health effects.
- For highly variable populations, provides best control of the extreme values.
- Useful for skewed distributions.
- May be appropriate when the population contains a large number of values less than the measurement detection limit, as long as this limit is less than the action level.
- Relies on few statistical assumptions.

Negative Attributes

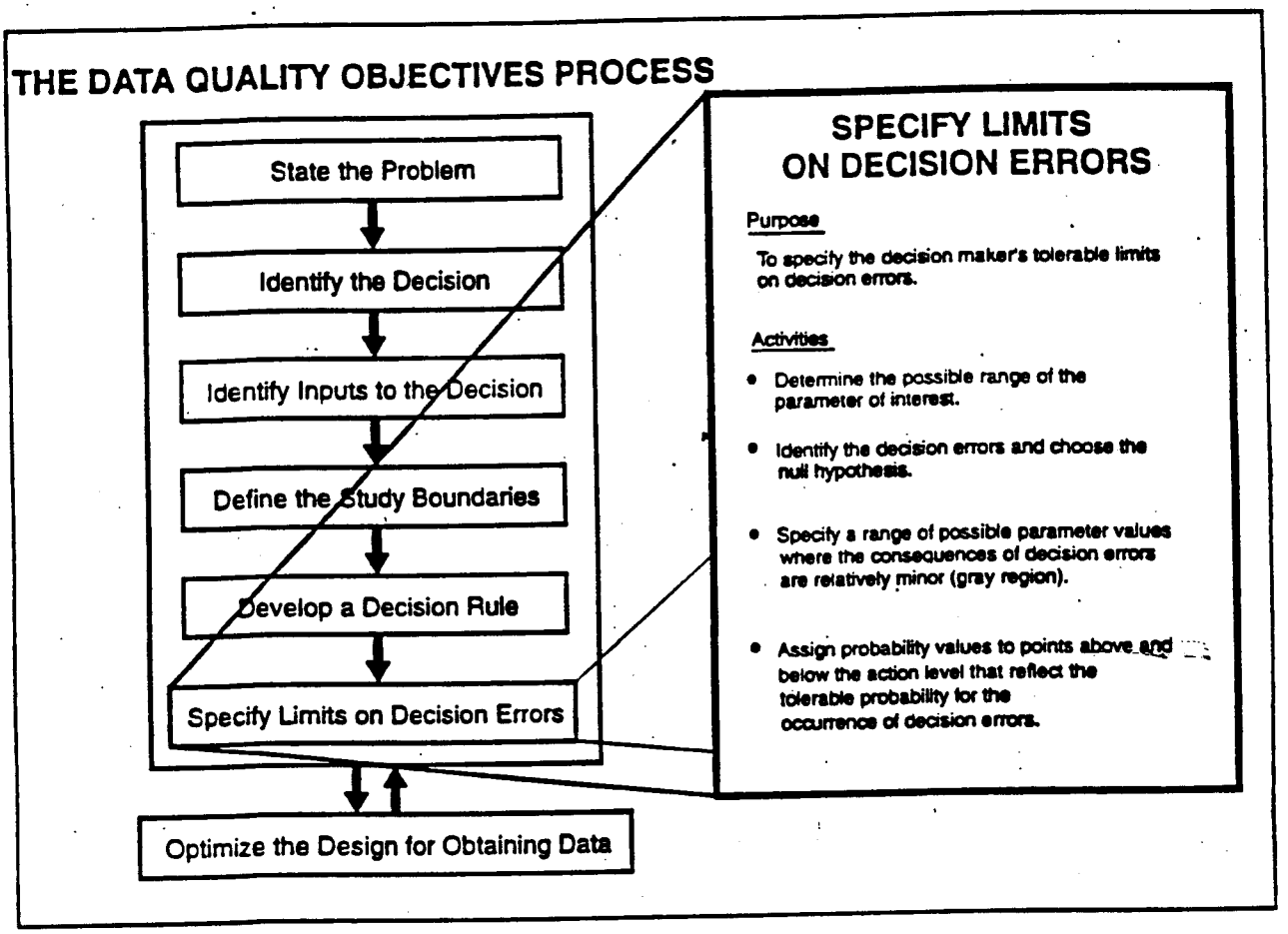
- Requires larger sample sizes than mean.

Reference: U.S. Environmental Protection Agency. 1989. *Methods for Evaluation Attainment of Cleanup Standards: Volume 1: Soils and Solid Media.* EPA 230/02-89-042. Office of Policy Planning and Evaluation.

Table 5-1. (cont.) Attributes of Different Statistical Parameters to Characterize the Population

CHAPTER 6

STEP 6: SPECIFY TOLERABLE LIMITS ON DECISION ERRORS



Purpose

The purpose of this step is to specify the decision maker's tolerable limits on decision errors, which are used to establish performance goals for the data collection design.

Expected Outputs

- **The decision maker's tolerable decision error rates based on a consideration of the consequences of making an incorrect decision.**

Background

Decision makers are interested in knowing the true state of some feature of the environment. Since data can only estimate this state, decisions that are based on measurement data could be in error (decision error). Most of the time the correct decision will be made; however, this chapter will focus on controlling the less likely possibility of making a decision error. The goal of the planning team is to develop a data collection design that reduces the chance of making a decision error to a tolerable level. This step of the DQO Process will provide a mechanism for allowing the decision maker to define tolerable limits on the probability of making a decision error.

There are two reasons why the decision maker cannot know the true value of a population parameter (i.e., the true state of some feature of the environment):

(1) The population of interest almost always varies over time and space. Limited sampling will miss some features of this natural variation because it is usually impossible or impractical to measure every point of a population. *Sampling design error* occurs when the sampling design is unable to capture the complete extent of natural variability that exists in the true state of the environment.

(2) Analytical methods and instruments are never absolutely perfect, hence a measurement can only estimate the true value of an environmental sample. *Measurement error* refers to a combination of random and systematic errors that inevitably arise during the various steps of the measurement process (for example, sample collection, sample handling, sample preparation, sample analysis, data reduction, and data handling).

The combination of sampling design error and measurement error is called *total study error*, which may lead to a decision error. Since it is impossible to eliminate error in measurement data, basing decisions on measurement data will lead to the possibility of making a decision error.

The probability of decision errors can be controlled by adopting a scientific approach. In this approach, the data are used to select between one condition of the environment (the *null hypothesis*, H_0) and an alternative condition (the *alternative hypothesis*, H_a). The null hypothesis is treated like a baseline condition that is presumed to be true in the absence of strong evidence to the contrary. This feature provides a way to guard against making the decision error that the decision maker considers to have the more undesirable consequences.

A decision error occurs when the decision maker rejects the null hypothesis when it is true, or fails to reject the null hypothesis when it is false. These two types of decision errors are classified as *false positive* and *false negative* decision errors, respectively. They are described below.

False Positive Decision Error — A false positive decision error occurs when the null hypothesis (H_0) is rejected when it is true. Consider an example where the decision maker presumes that a certain waste is hazardous (i.e., the null hypothesis or baseline condition is "the waste is hazardous"). If the decision maker concludes that there is insufficient evidence to classify the waste as hazardous when it truly is hazardous, then the decision maker would make a false positive decision error. A statistician usually refers to the false positive error as a "Type I" error. The measure of the size of this error is called alpha (α), the level of significance, or the size of the critical region.

False Negative Decision Error — A false negative decision error occurs when the null hypothesis is <u>not</u> rejected when it is false. In the above waste example, the false negative decision error occurs when the decision maker concludes that the waste is hazardous when it truly is <u>not</u> hazardous. A statistician usually refers to a false negative error as a "Type II" error. The measure of the size of this error is called beta (β), and is also known as the complement of the *power* of a hypothesis test.

The definition of false positive and false negative decision errors depends on the viewpoint of the decision maker.[1] Consider the viewpoint where a person has been presumed to be "innocent until proven guilty" (i.e., H_0 is "innocent"; H_a is "guilty"). A false positive error would be convicting an innocent person; a false negative error would be not convicting the guilty person. From the viewpoint where a person is presumed to be "guilty until proven innocent" (i.e., H_0 is "guilty"; H_a is "innocent"), the errors are reversed. Here, the false positive error would be not convicting the guilty person, and the false negative error would be convicting the innocent person.

While the possibility of a decision error can never be totally eliminated, it can be controlled. To control the possibility of making decision errors, the planning team must control total study error. There are many ways to accomplish this, including collecting a large number of samples (to control sampling design error), analyzing individual samples several times or using more precise laboratory methods (to control measurement error). Better sampling designs can also be developed to collect data that more accurately and efficiently represent the population of interest. Every study will use a slightly different method of controlling decision errors, depending on where the largest components of total study error exist in the data set and the ease of reducing those error components. Reducing the probability of making decision errors generally increases costs. In many cases controlling decision error within very small limits is unnecessary for making a decision that satisfies the decision maker's needs. For instance, if the consequences of decision errors are minor, a reasonable decision could be made based on relatively crude data (data with high total study

[1]Note that these definitions are not the same as false positive or false negative instrument readings, where similar terms are commonly used by laboratory or field personnel to describe a fault in a single result; false positive and false negative <u>decision</u> errors are defined in the context of hypothesis testing, where the terms are defined with respect to the null hypothesis.

error). On the other hand, if the consequences of decision errors are severe, the decision maker will want to control sampling design and measurement errors within very small limits.

To minimize unnecessary effort controlling decision errors, the planning team must determine whether reducing sampling design and measurement errors is necessary to meet the decision maker's needs. These needs are made explicit when the decision maker specifies probabilities of decision errors that are tolerable. Once these tolerable limits on decision errors are defined, then the effort necessary to analyze and reduce sampling design and measurement errors to satisfy these limits can be determined in Step 7: Optimize the Design for Obtaining Data. It may be necessary to iterate between these two steps before finding tolerable probabilities of decision errors that are feasible given resource constraints.

Activities

Determine the possible range of the parameter of interest. Establish the possible range of the parameter of interest by estimating its likely upper and lower bounds. This will help focus the remaining activities of this step on only the relevant values of the parameter. For example, the range of the parameter shown in Figures 6-1 and 6-2 at the end of this chapter is between 50 and 200 ppm. Historical and documented analytical data are of great help in establishing the potential parameter range.

Identify the decision errors and choose the null hypothesis. Define where each decision error occurs relative to the action level and establish which decision error should be defined as the null hypothesis (baseline condition). This process has four steps:

(1) *Define both types of decision errors and establish the true state of nature for each decision error.* Define both types of decision errors and determine which one occurs above and which one occurs below the action level. A decision error occurs when the data mislead the decision maker into concluding that the parameter of interest is on one side of the action level when the true value of the parameter is on the other side of the action level. For example, consider a situation in which a study is being conducted to determine if mercury contamination is creating a health hazard and EPA wants to take action if more than 5% of a population of fish have mercury levels above a risk-based action level. In this case, a decision error would occur if the data lead the decision maker to conclude that 95% of the mercury levels found in the fish population were below the action level (i.e., the parameter is the "95th percentile" of mercury levels in the fish population) when the true 95th percentile of mercury levels in the fish population was above the action level (which means that more than 5% of the fish population contain mercury levels greater than the action level). The other decision error for this example would be that the data lead the decision maker to conclude that the 95th percentile of mercury levels in the fish population is greater than the action level when the true 95th percentile is less than the action level.

The "true state of nature" is the actual condition or feature of the environment that exists, but is unknown to the decision maker. Each decision error consists of two parts, the true state of nature and the conclusion that the decision maker draws. Using the example above, the true state of nature for the first decision error is that the 95th percentile of mercury levels in the fish population is above the action level.

(2) *Specify and evaluate the potential consequences of each decision error.* Specify the likely consequences of making each decision error and evaluate their potential severity in terms of economic and social costs, human health and ecological effects, political and legal ramifications, and so on. Consider the alternative actions that would be taken under each decision error scenario, as well as secondary effects of those actions. For example, in determining whether or not 95% of a fish population contain mercury levels above a risk-based action level, there may be a variety of potential consequences of committing a decision error. In the first decision error described above, where the decision maker concludes that the 95th percentile is below when the true 95th percentile was above the action level, the decision maker may decide to continue to allow fishing in the waters and not undertake any cleanup activity. The resulting consequences might include human health and ecological effects from consumption of contaminated fish by humans and other animals, economic and social costs of health care and family disruption, and damaged credibility of EPA when (and if) the decision error is detected. If the other type of decision error is committed, where the decision maker decides that the 95th percentile exceeds the action level when the true 95th percentile is below the action level, the decision maker might ban all fishing in the local waters and initiate cleanup activities. The consequences might include economic and social costs of lost revenues and job displacement in the fishing industry, damaged credibility for EPA when the cleanup activities expose the nature of the decision error, and the threat of lawsuits by fishing interests.

Evaluate the severity of potential consequences of decision errors at different points within the domains of each type of decision error, since the severity of consequences may change as the parameter moves further away from the action level. Consider whether or not the consequences change abruptly at some value, such as a threshold health effect level; the decision maker may want to change the tolerable limit on the decision error at such a point.

(3) *Establish which decision error has more severe consequences near the action level.* Based on the evaluation of potential consequences of decision errors, the decision maker should determine which decision error causes greater concern when the true parameter value is near the action level. It is important to focus on the region near the action level because this is where the true parameter value is most likely to be when a decision error is made (in other words, when

the true parameter is far above or far below the action level, the data are much more likely to indicate the correct decision). This determination typically involves value judgements about the relative severity of different types of consequences within the context of the problem. In the fish contamination problem above, the decision maker would weigh the potential health consequences from allowing people to consume contaminated fish versus the economic and social disruption from banning all fishing in the community. In this case, the decision maker might carefully consider how uncertain or conservative the risk-based action level is.

(4) *Define the null hypothesis (baseline condition) and the alternative hypothesis and assign the terms "false positive" and "false negative" to the appropriate decision error.* In problems that concern regulatory compliance, human health, or ecological risk, the decision error that has the most adverse potential consequences should be defined as the null hypothesis (baseline condition).[2] In statistical hypothesis testing, the data must conclusively demonstrate that the null hypothesis is false. That is, the data must provide enough information to authoritatively reject the null hypothesis (disprove the baseline condition) in favor of the alternative. Therefore, by setting the null hypothesis equal to the true state of nature that exists when the more severe decision error occurs, the decision maker guards against making the more severe decision error by placing the burden of proof on demonstrating that the most adverse consequences will <u>not</u> be likely to occur.

It should be noted that the null and alternative hypotheses have been predetermined in many regulations. If not, the planning team should define the null hypothesis (baseline condition) to correspond to the true state of nature for the more severe decision error and define the alternative hypothesis to correspond to the true state of nature for the less severe decision error.

Using the definitions of null and alternative hypotheses, assign the term "false positive" to the decision error in which the decision maker rejects the null hypothesis when it is true, which corresponds to the decision error with the more severe consequences identified in task (3). Assign the term "false negative" to the decision error in which the decision maker fails to reject the

---

[2]Note that this differs somewhat from the conventional use of hypothesis testing in the context of planned experiments. There, the alternative hypothesis usually corresponds to what the experimenter hopes to prove, and the null hypothesis usually corresponds to some baseline condition that represents an "opposite" assumption. For instance, the experimenter may wish to prove that a new water treatment method works better than an existing accepted method. The experimenter might formulate the null hypothesis to correspond to "the new method performs no better than the accepted method," and the alternative hypothesis as "the new method performs better than the accepted method." The burden of proof would then be on the experimental data to show that the new method performs better than the accepted method, and that this result is not due to chance.

null hypothesis when it is false, which corresponds to the decision error with the less severe consequences identified in task (3).

Specify a range of possible parameter values where the consequences of decision errors are relatively minor (gray region). The gray region is a range of possible parameter values where the consequences of a false negative decision error are relatively minor. The gray region is bounded on one side by the action level and on the other side by that parameter value where the consequences of making a false negative decision error begin to be significant. Establish this boundary by evaluating the consequences of not rejecting the null hypothesis when it is false. The edge of the gray region should be placed where these consequences are severe enough to set a limit on the magnitude of this false negative decision error. Thus, the gray region is the area between this parameter value and the action level.

It is necessary to specify a gray region because variability in the population and unavoidable imprecision in the measurement system combine to produce variability in the data such that a decision may be "too close to call" when the true parameter value is very near the action level. Thus, the gray region (or "area of uncertainty") establishes the minimum distance from the action level where the decision maker would like to begin to control false negative decision errors. In statistics, the width of this interval is called the "minimum detectable difference" and is often expressed as the Greek letter delta (Δ). The width of the gray region is an essential part of the calculations for determining the number of samples needed to satisfy the DQOs, and represents one important aspect of the decision maker's concern for decision errors. A more narrow gray region implies a desire to detect conclusively the condition when the true parameter value is close to the action level ("close" relative to the variability in the data). When the true value of the parameter falls within the gray region, the decision maker may face a high probability of making a false negative decision error, since the data may not provide conclusive evidence for rejecting the null hypothesis, even though it is actually false (i.e., the data may be too variable to allow the decision maker to recognize that the presumed baseline condition is, in fact, <u>not</u> true).

From a practical standpoint, the gray region is an area where it will not be feasible or reasonable to control the false negative decision error rate to low levels because of high costs. Given the resources that would be required to reliably detect small differences between the action level and the true parameter value, the decision maker must balance the resources spent on data collection with the expected consequences of making that decision error. For example, when testing whether a parameter (such as the mean concentration) exceeds the action level, if the <u>true</u> parameter is near the action level (relative to the expected variability of the data), then the imperfect data will tend to be clustered around the action level, with some values above the action level and some below. In this situation, the likelihood of committing a false negative decision error will be large. To determine with confidence whether the true value of the parameter is above or below the action level, the decision maker would need to collect a large amount of data, increase the precision of the measurements, or both. If taken to an extreme, the cost of collecting data can exceed the cost of making a decision error, especially where the consequences of the decision error may be relatively

minor. Therefore, the decision maker should establish the gray region, or the region where it is not critical to control the false negative decision error, by balancing the resources needed to "make a close call" versus the consequences of making that decision error.

Assign probability limits to points above and below the gray region that reflect the tolerable probability for the occurrence of decision errors. Assign probability values to points above and below the gray region that reflect the decision maker's tolerable limits for making an incorrect decision. Select a possible value of the parameter; then choose a probability limit based on an evaluation of the seriousness of the potential consequences of making the decision error if the true parameter value is located at that point. At a minimum, the decision maker should specify a false positive decision error limit at the action level, and a false negative decision error limit at the other end of the gray region. For many situations, the decision maker may wish to specify additional probability limits at other possible parameter values. For example, consider a hypothetical toxic substance that has a regulatory action level of 10 ppm, and which produces threshold effects in humans exposed to mean concentrations above 100 ppm. In this situation, the decision maker may wish to specify more stringent probability limits at that threshold concentration of 100 ppm than those specified at 10 ppm. The tolerable decision error limits should decrease further away from the action level as the consequences of decision error become more severe.

Given the potentially high cost of controlling sampling design error and measurement error for environmental data, Agency decision making is rarely supported by decision error limits more stringent than 0.01 (1%) for both the false positive and false negative decision errors. This guidance recommends using 0.01 as the starting point for setting decision error rates. The most frequent reasons for setting limits greater (i.e., less stringent) than 0.01 are that the consequences of the decision errors may not be severe enough to warrant setting decision error rates that are this extreme. The value of 0.01 should <u>not</u> be considered a prescriptive value for setting decision error rates, nor should it be considered as the policy of EPA to encourage the use of any particular decision error rate. Rather, it should be viewed as a starting point from which to develop limits on decision errors that are applicable for each study. If the decision maker chooses to relax the decision error rates from 0.01 for false positive or false negative decision errors, the planning team should document the reasoning behind setting the less stringent decision error rate and the potential impacts on cost, resource expenditure, human health, and ecological conditions.

The combined information from the activities section of this chapter can be graphed onto a "Decision Performance Goal Diagram" or charted in a "Decision Error Limits Table" (see Figures 6-1 and 6-2 and Tables 6-1 and 6-2 below). Both are useful tools for visualizing and evaluating all of the outputs from this step. Figure 6-1 and Table 6-1 illustrate the case where the null hypothesis (baseline condition) is that the parameter of interest exceeds the action level (e.g., the waste is hazardous). Figure 6-2 and Table 6-2 illustrate the case where the null hypothesis (baseline condition) is that the parameter is less than the action level (e.g., the waste is not hazardous).

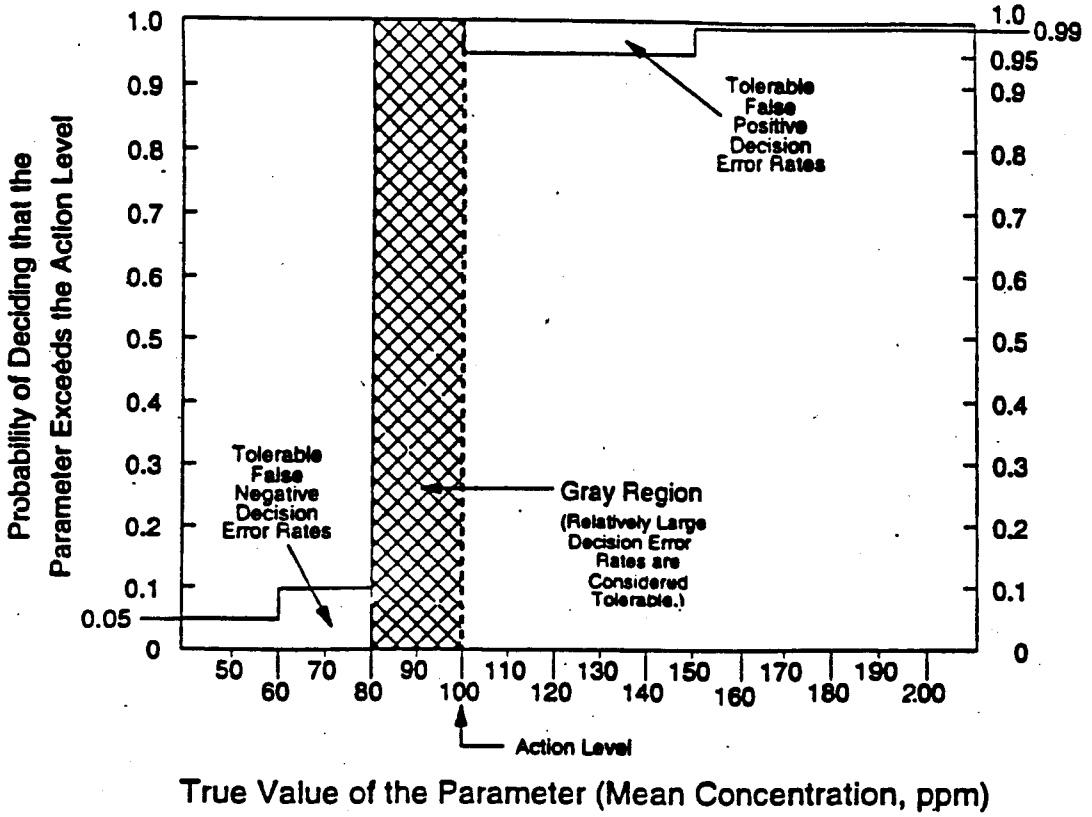


Figure 6-1. An Example of a Decision Performance Goal Diagram Baseline Condition: Parameter Exceeds Action Level.

True Concentration	Correct Decision	Type of Error	Tolerable Probability of Incorrect Decision
< 60 ppm	Not exceed	F(-)	5%
60 to 80	Not exceed	F(-)	10%
80 to 100	Not exceed	F(-)	gray region
100 to 150	Does exceed	F(+)	5%
> 150	Does exceed	F(+)	1%

Table 6-1. Decision Error Limits Table Corresponding to Figure 6-1. (Action Level = 100 ppm)

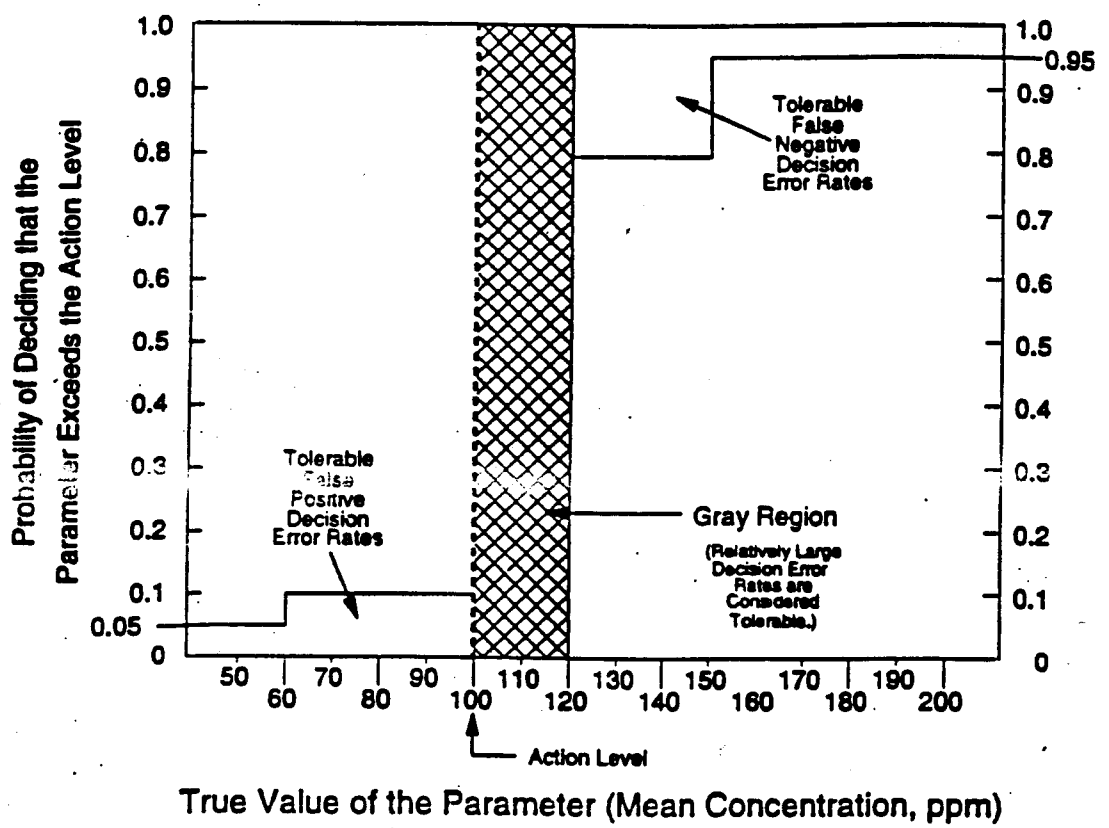


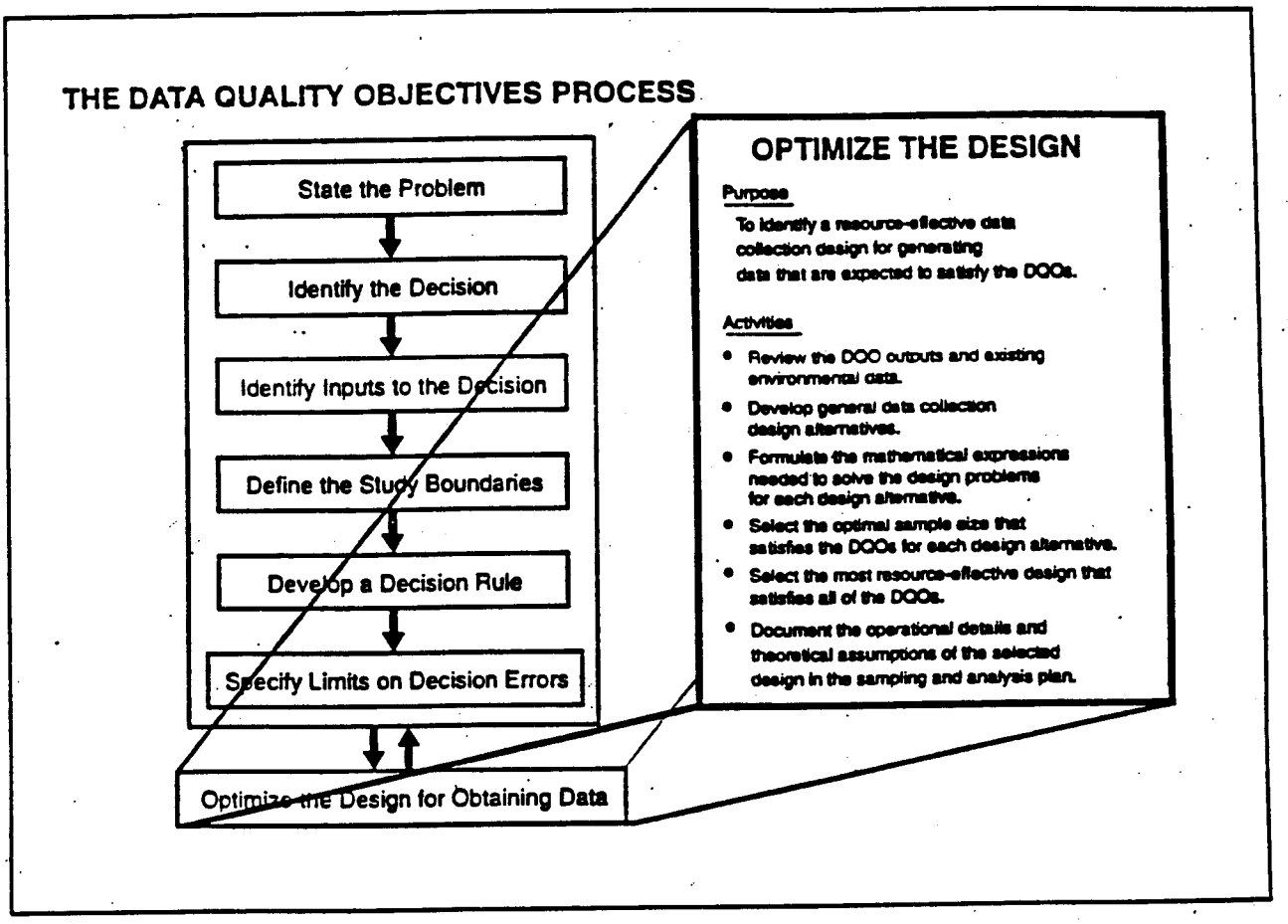
Figure 6-2. An Example of a Decision Performance Goal Diagram Baseline Condition: Parameter Is Less Than Action Level.

True Concentration	Correct Decision	Type of Error	Tolerable Probability of Incorrect Decision
< 60 ppm	Not exceed	F(+)	5%
60 to 100	Not exceed	F(+)	10%
100 to 120	Does exceed	F(-)	gray region
120 to 150	Does exceed	F(-)	20%
> 150	Does exceed	F(-)	5%

Table 6-2. Decision Error Limits Table Corresponding to Figure 6-2.

CHAPTER 7

STEP 7: OPTIMIZE THE DESIGN FOR OBTAINING DATA



Purpose

The purpose of this step is to identify a resource-effective data collection design for generating data that are expected to satisfy the DQOs.

Expected Outputs

- **The most resource-effective design for the study that is expected to achieve the DQOs.**

Background

In this step, statistical techniques are used to develop alternative data collection designs and evaluate their efficiency in meeting the DQOs. To develop the optimal design for this study, it may be necessary to work through this step more than once after revisiting previous steps of the DQO Process.

The objective of this step is to identify the most resource-effective data collection design expected to generate data that satisfy the DQOs specified in the preceding steps. While a full explanation of the procedures for developing a data collection design is beyond the scope of this guidance document, it does provide a broad overview of the steps that need to be accomplished to reach this goal. The example in Appendix B illustrates some of these activities in more detail.

Activities

Review the DQO outputs and existing environmental data. Review the DQO outputs generated in the preceding six steps to ensure that they are internally consistent. The DQOs should provide a succinct collection of information on the context of, requirements for, and constraints on the data collection design. 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 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.

Develop general data collection design alternatives. 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. Generally, the goal is to find cost-effective alternatives that balance sample size and measurement performance, given the feasible choices for sample collection techniques and analytical methods. In some cases where there is a relatively high spatial or temporal variability, it may be more cost-effective to use less expensive yet less precise analytical methods so that a relatively large number of samples can be taken, thereby controlling the sampling design error component of total study error. In other cases where the contaminant distribution is relatively homogeneous, or the action level is very near the method detection limit, it may be more cost-effective to use more expensive yet more precise and/or more sensitive analytical methods and collect fewer samples, thereby controlling the analytical measurement error component of total study error. Examples of general data collection design alternatives include:

- factorial design
- simple random sampling
- stratified random sampling
- sequential random sampling
- systematic sampling
- composite sampling (in conjunction with another sampling design)

Formulate the mathematical expressions needed to solve the design problem for each data collection design alternative. Develop the following three mathematical expressions needed to optimize the data collection design as follows:

(1) Define a suggested method for testing the statistical hypothesis and define a sample size formula that corresponds to the method if one exists (e.g., a Student's t-test).

(2) Develop a statistical model that describes the relationship of the measured value to the "true" value. Often the model will describe the components of error or bias that are believed to exist in the measured value.

(3) Develop a cost function that relates the number of samples to the total cost of sampling and analysis.

Select the optimal sample size that satisfies the DQOs for each data collection design alternative. Using the mathematical expressions from the previous activity, solve for the optimal sample size that satisfies the DQOs, including the decision maker's limits on decision errors. If no design will meet the limits on decision errors within the budget or other constraints, then the planning team will need to relax one or more constraints. For example:

- increase the budget for sampling and analysis;
- increase the width of the gray region;
- increase the tolerable decision error rates;
- relax other project constraints, such as the schedule; or
- change the boundaries; it may be possible to reduce sampling and analysis costs by changing or eliminating subgroups that will require separate decisions.

Select the most resource-effective data collection design that satisfies all of the DQOs. Evaluate the design options based on cost and ability to meet the DQO constraints. Choose the one that provides the best balance between cost (or expected cost) and ability to meet the DQOs.

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.[1] The performance curve can be overlaid into

[1]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.

the Decision Performance Goal Diagram to assess how well a test performs or to compare competing tests. A design that produces a very steep performance curve is preferred over one that is relatively flat. An example of a performance curve is shown in Figure 7-1.

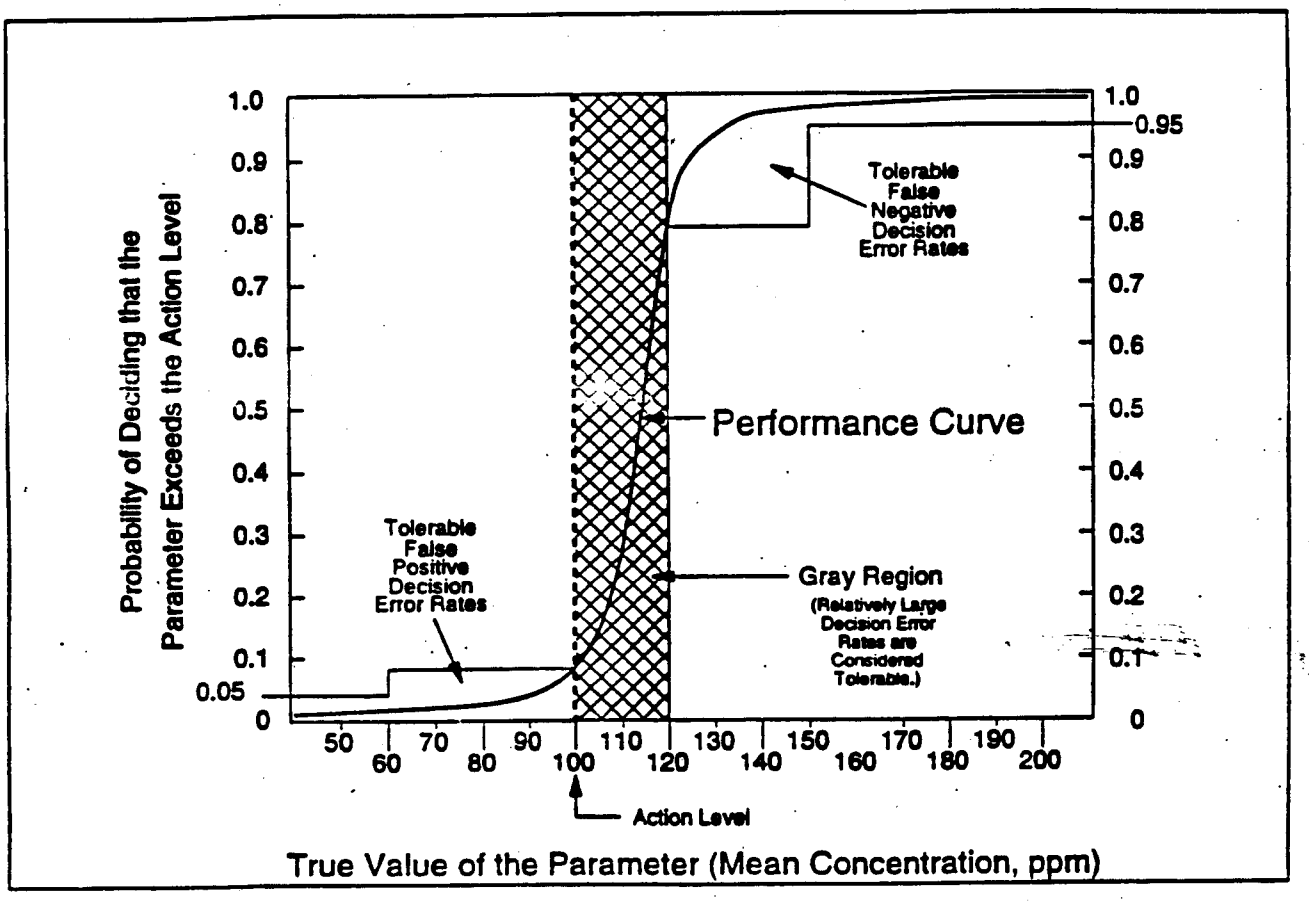


Figure 7-1. An Example of a Power Curve
Baseline Condition: Parameter is Less Than Action Level.

Document the operational details and theoretical assumptions of the selected design in the sampling and analysis plan. Document the selected design's key features that must be implemented properly to allow for efficient and valid statistical interpretation of the data. It is particularly important to document the statistical assumptions that could be violated through errors in or practical constraints on field sample collection procedures or analytical methods.

After all the activities have been completed it may be helpful to enlist the advice and review of a statistician with expertise in data collection designs. This will be particularly useful if the initial data collection designs have been developed by an inexperienced statistician or an environmental scientist with limited statistical training. The experienced statistician may be able to offer innovative alternative data collection designs that may be more cost-effective or simpler to implement.

BIBLIOGRAPHY

Bates, D.J., R.O. Gilbert, N.L. Hassig, R.F. O'Brien, B.A. Pulsipher. November 1993. Decision Performance Criteria: The Driver Behind the Data Quality Objectives Process, A Statistical Introduction (Draft). Battelle Pacific Northwest Laboratory, Richland, Washington.

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

Desu, M.M., and D. Raghavarao. 1990. *Sample Size Methodology.* San Diego, CA: Academic Press.

Gilbert, Richard O. 1987. *Statistical Methods for Environmental Pollution Monitoring.* New York: Von Nostrand Reinhold.

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.* Vol. 35, No. 4.

U.S. Environmental Protection Agency. 1994. *EPA Quality System Requirements for Environmental Programs.* EPA QA/R-1.

U.S. Environmental Protection Agency. 1994. *EPA Requirements for Quality Assurance Project Plans for Environmental Data Operations.* EPA QA/R-5.

U.S. Environmental Protection Agency. 1994. *EPA Requirements for Quality Management Plans.* EPA QA/R-2.

U.S. Environmental Protection Agency. 1994. *Guidance for Data Quality Assessments.* EPA QA/G-9.

U.S. Environmental Protection Agency. 1993. *Guidance for Planning in Support of Environmental Decision Making Using the Data Quality Objectives Process* (Interim Final). Quality Assurance Management Staff.

U.S. Environmental Protection Agency. 1992. *Statistical Methods for Evaluating the Attainment of Cleanup Standards: Volume III: Reference-Based Standards for Soils and Solid Media.* EPA 230-R-94-004, Office of Policy, Planning and Evaluation.

U.S. Environmental Protection Agency. 1992. *Methods for Evaluating the Attainment of Cleanup Standards: Volume 2: Ground Water.* EPA 230-R-92-014, Office of Policy, Planning and Evaluation.

U.S. Environmental Protection Agency. 1989. *Methods for Evaluating Attainment of Cleanup Standards: Volume 1: Soils and Solid Media.* EPA 230/02-89-042, Office of Policy, Planning and Evaluation.

U.S. Environmental Protection Agency. 1986. *Development of Data Quality Objectives, Description of Stages I and II.* Quality Assurance Management Staff.

U.S. Environmental Protection Agency. April 1984. "Order 5360.1, Policy and Program Requirements to Implement the Mandatory Quality Assurance Program." Office of the Administrator.

APPENDIX A

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

Overview

This appendix explains some important QA management steps that occur after the DQO Process has been completed. The DQO Process is part of the planning phase of the data collection operation, as illustrated in Figure A-1. At the completion of the DQO Process, the planning team will have documented the project objectives and key performance requirements for the data operations in the DQOs, and will have identified a data collection design that is expected to achieve the DQOs. The data collection design and DQOs will then be used to develop the Quality Assurance Project Plan (QAPP), which provides the detailed project-specific objectives, specifications, and procedures needed to conduct a successful data collection activity. During the implementation phase of the data collection life cycle, the QAPP is executed and the data are collected. During the assessment phase, a Data Quality Assessment (DQA) is performed on the data to determine if the DQOs have been satisfied. The relationship between the DQO Process and these subsequent activities are explained in more detail below.

Quality Assurance Project Plan Development

The QAPP is a formal EPA project document that specifies the operational procedures and quality assurance/quality control (QA/QC) requirements for obtaining environmental data of sufficient quantity and quality to satisfy the project objectives. The QAPP is an important part of the EPA Quality System, and is required for all data collection activities that generate data for use by EPA.[1] The QAPP contains information on project management, measurement and data acquisition, assessment and oversight, and data validation and useability.

The DQO Process may be viewed as a preliminary step in the QAPP development process, as shown in the right half of Figure A-1. DQOs are a formal element of the QAPP, yet information contained in the DQOs relates indirectly to many other elements of the QAPP. In essence, the DQOs provide statements about the expectations and requirements of the data *user* (such as a decision maker). In the QAPP, these requirements are translated into measurement performance specifications and QA/QC procedures for the data *suppliers*, to

[1]U.S. Environmental Protection Agency. *EPA Requirements for Quality Assurance Project Plans for Environmental Data Operations.* EPA QA/R-5, 1994.

provide them with the information they need to satisfy the data user's needs. Thus, the QAPP integrates the DQOs, the data collection design, and QA/QC procedures into a coherent plan to be used for collecting defensible data that are of known quality and that is adequate for the data's intended use.

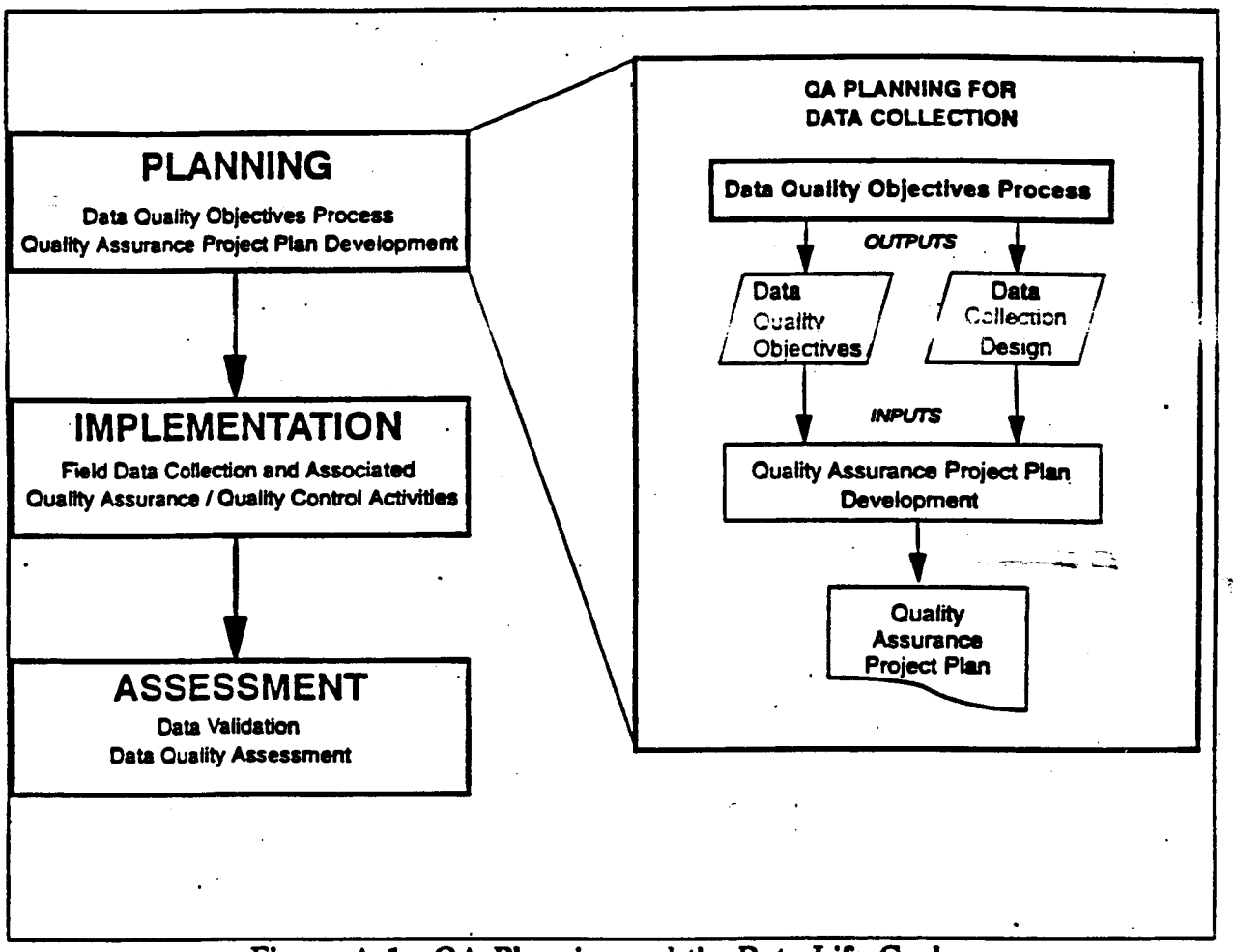


Figure A-1. QA Planning and the Data Life Cycle.

The QAPP is structured into three sections: the Introduction, Requirements, and Elements. The Elements are the individual requirements of the QAPP that are listed separately. The Elements are grouped into four categories: Project Management, Measurement/Data Acquisition, Assessment/Oversight, and Data Validation and Useability. The outputs of the DQO Process will provide information or inputs to elements in the Project Management section.

Data Quality Assessment

After the environmental data have been collected and validated in accordance with the QAPP, the data must be evaluated to determine whether the DQOs have been satisfied. EPA has developed guidance on Data Quality Assessment (DQA) to address this need (see Figure A-2).[2] DQA involves the application of statistical tools to determine:

- whether the data meet the assumptions under which the DQOs and the data collection design were developed; and
- whether the total error in the data is small enough to allow the decision maker to use the data to support the decision within the tolerable decision error rates expressed by the decision maker.

It is important to verify the assumptions that underlie the DQOs and the data collection design so that statistical calculations performed on the data relate to the decision maker's problem in a scientifically valid and meaningful way. If the data do not support the underlying assumptions, then corrective actions must be taken to ensure that the decision maker's needs are met. Corrective action may be as simple as selecting a different statistical approach that relies on assumptions that are in better agreement with the data, or it may be as complicated as revising the data collection design and collecting new data that satisfy the decision maker's needs.

If the data support the conclusion that the assumptions are reasonable, then the next step of a DQA can be taken, which is to evaluate how well the data support the actual decision. This is determined by evaluating whether the data conclusively demonstrate that the population parameter of interest is above (or below) the action level. In essence, this is where the decision maker applies a more specific or "operational" version of the decision rule that was developed in Step 5 of the DQO Process (in statistical terms, this is performing the hypothesis test). Whether the data are "conclusive" or not will depend on the estimated value and variability of the statistical parameter in relation to the gray region and the limits on decision errors that were specified in Step 6 of the DQO Process. If the decision cannot be made in accordance with the decision maker's DQOs, then the decision maker must decide whether to take corrective actions (such as collect more or better data), relax the DQOs, or make a decision anyway, without the benefit of adequate data.

[2] U. S. Environmental Protection Agency. *Guidance for Data Quality Assessments.* EPA QA/G-9, 1994.

Thus, DQA is an essential element of the data operation because it helps to bring closure to the issues raised at the beginning of the DQO Process. By verifying the assumptions required to draw scientifically valid and meaningful conclusions from the data, and by implementing the decision rule, DQA helps the decision maker determine whether the DQOs have been satisfied.

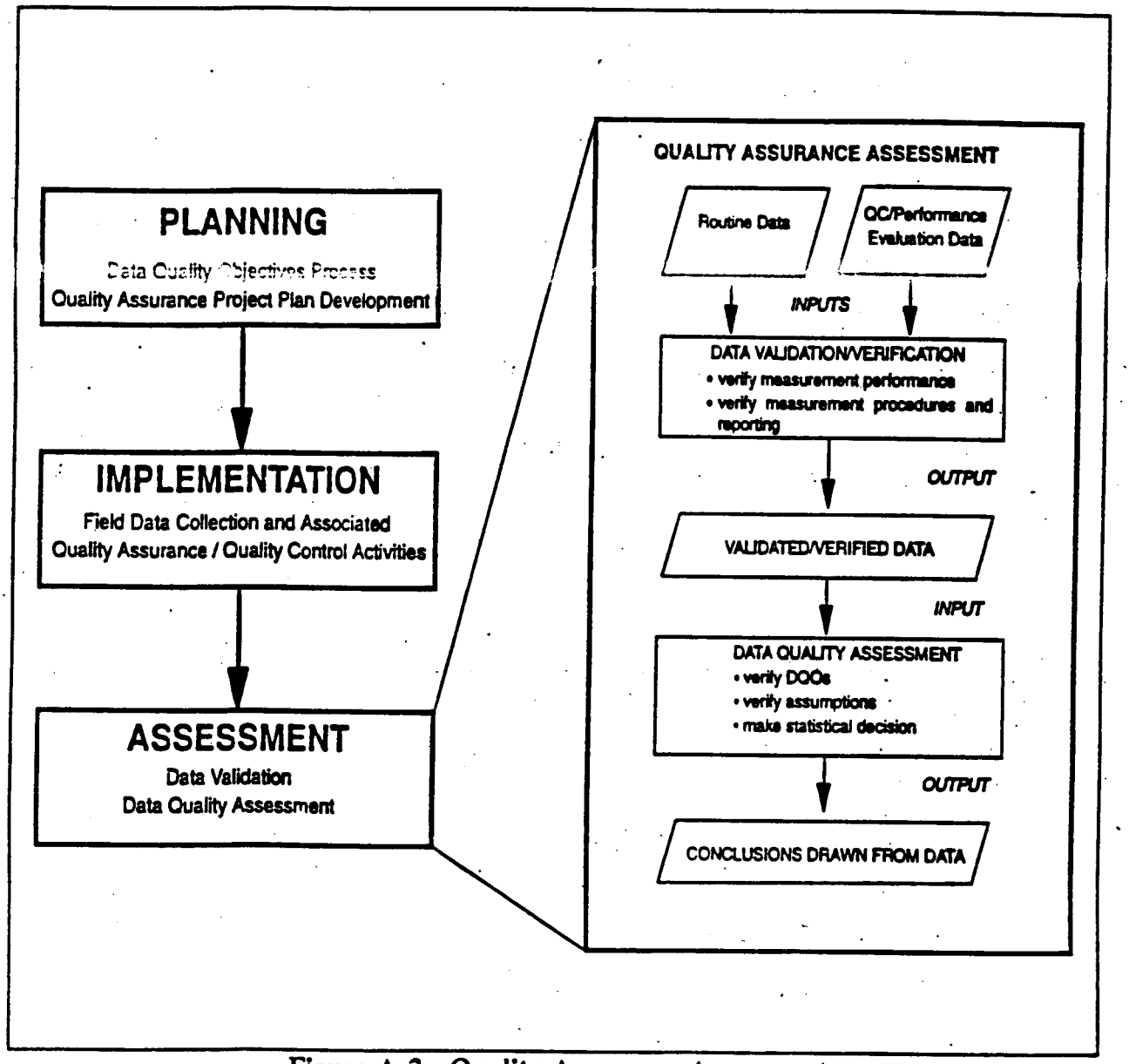


Figure A-2. Quality Assurance Assessment.

APPENDIX B

DQO CASE STUDY: CADMIUM-CONTAMINATED FLY ASH WASTE

Introduction

This appendix presents a functional, but realistic example of the DQO outputs for a decision that could be made within the Resource Conservation and Recovery Act (RCRA) hazardous waste management program. The example is intended to illustrate the types of outputs that are common to the DQO Process. It is not intended, however, to represent the policy of the RCRA program for actual situations that may be similar to the example. Please consult with a knowledgeable representative within the RCRA program office about the current policy for making waste classification decisions for fly ash or other types of hazardous waste.

The case study has been chosen because it is simple and straightforward, and because the outputs are uncomplicated. Although some of the outputs from this example may seem intuitive, this is not often the case in practice. For many studies, the DQO Process is complicated and thought-provoking. Even so, some steps will require more effort than others. Keep in mind that all of the steps in the DQO Process are necessary to develop a data collection design. Once the first six steps have been completed and thoroughly thought-out, then development of the most resource-effective data collection design can proceed.

Background

A waste incineration facility located in the Midwest routinely removes fly ash from its flue gas scrubber system and disposes of it in a local sanitary landfill. Previously it was determined that the ash was not hazardous according to RCRA program regulations. The incinerator, however, recently began treating a new waste stream. The representatives of the incineration company are concerned that the waste fly ash could now contain hazardous levels of cadmium from the 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. They have decided to employ the DQO Process to help guide their decision making.

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

Under the current Code of Federal Regulations, 40 CFR, Part 261, a solid waste can be considered "hazardous" if it meets specific criteria of ignitability, corrosivity, reactivity, and toxicity. One method that is used to determine if a solid substance, such as fly ash, meets the criteria for toxicity under the RCRA program regulations is to test a "representative sample" of the waste and perform a Toxicity Characteristic Leaching Procedure (TCLP) described in 40 CFR, Pt. 261, App. II. During this process, the solid fly ash will be "extracted" using an acid solution. The extraction liquid (the TCLP leachate) will then be subjected to tests for specific metals and compounds. For this example, the only concern is with the concentration of cadmium in the leachate. The primary benefit of the DQO Process will be to establish the data collection design needed to determine if the waste is hazardous under RCRA regulations within tolerable decision error rates.

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 between loads of ash leaving the facility. They have determined that each load is fairly homogeneous. There is a high variability between loads, however, due to the nature of the waste stream. Most of the fly ash produced is not hazardous and may be disposed of in a sanitary landfill. Thus, the company has decided that testing each individual waste load before it leaves the facility would be the most economical. Then they could send loads of ash that exceeded the regulated standards to the higher-cost RCRA landfills and continue to send the others to the sanitary landfill.

DQO Development

The following is a representative example of the output from each step of the DQO Process for the fly ash toxicity problem.

State the Problem — a description of the problem(s) and specifications of available resources and relevant deadlines for the study.

(1) *Identify the members of the planning team* — The members of the planning team will include the incineration plant manager, a plant engineer, a statistician, a quality assurance officer, an EPA representative who works within the RCRA program, and a chemist with sampling experience.

(2) *Identify the primary decision maker* — There will not be a primary decision maker; decisions will be made by consensus.

(3) *Develop a concise description of the problem* — The problem is to determine which loads should be sent to a RCRA landfill versus a sanitary landfill.

(4) *Specify available resources and relevant deadlines for the study* — While the project will not by constrained by cost, the waste generator (the incineration company) wishes to hold sampling costs below $2,500. They have also requested that the waste testing be completed within 1 week for each container load.

Identify the Decision — a statement of the decision that will use environmental data and the actions that could result from this decision.

(1) *Identify the principal study question* — Is the fly ash waste considered hazardous under RCRA regulations?

(2) *Define alternative actions that could result from resolution of the principal study question* —

(a) The waste fly ash could be disposed of in a RCRA landfill.

(b) The waste fly ash could be disposed of in a sanitary landfill.

(3) *Combine the principal study question and the alternative actions into a decision statement* — Decide whether or not the fly ash waste is hazardous under RCRA and requires special disposal procedures.

(4) *Organize multiple decisions* — Only one decision is being evaluated.

Identify the Inputs to the Decision — a list of the environmental variables or characteristics that will be measured and other information needed to resolve the decision statement.

(1) *Identify the information that will be required to resolve the decision statement* — To resolve the decision statement, the planning team needs to obtain measurements of the cadmium concentration in the leachate resulting from TCLP extraction.

(2) *Determine the sources for each item of information identified* — The fly ash should be tested to determine if it meets RCRA regulated standards for toxicity using the test methods listed in 40 CFR, Pt. 261, App. II. Existing pilot study data provide information about variability, but do not provide enough information to resolve the decision statement.

(3) *Identify the information that is needed to establish the action level* — The action level will be based on the RCRA regulations for cadmium in TCLP leachate.

(4) *Confirm that appropriate measurement methods exist to provide the necessary data* — Cadmium can be measured in the leachate according to the method specified in 40 CFR, Pt. 261, App. II. The detection limit is below the standard.

Define the Boundaries of the Study — a detailed description of the spatial and temporal boundaries of the problem, characteristics that define the population of interest, and any practical considerations for the study.

(1) *Specify the characteristics that define the population of interest* — Fly ash waste from the hazardous waste incinerator will be analyzed. The fly ash should not be mixed with any other constituents except water that is used for dust control. Each load of ash should fill at least 70% of the waste trailer. In cases where the trailer is filled less than 70%, the trailer must wait on-site until more ash is produced and fills the trailer to the appropriate capacity.

(2) *Define the spatial boundary of the decision statement* —

(a) *Define the geographic area to which the decision statement applies.* Decisions will apply to each container load of fly ash waste.

(b) *When appropriate, divide the population into strata that have relatively homogeneous characteristics.* Stratification is not necessary since the waste ash is relatively homogeneous within each container.

(3) *Define the temporal boundary of the decision statement* —

(a) *Determine the timeframe to which the decision statement applies.* It will be assumed that the sampling data represent both the current and future concentration of cadmium within the ash.

(b) *Determine when to collect data.* Contained in the trucks, the waste does not pose a threat to humans or the environment. Additionally, since the fly ash is not subject to change, disintegration, or alteration, the decision about the waste characteristics does not warrant any temporal constraints. To expedite decision making, however, the planning team has placed deadlines on sampling and reporting. The fly ash waste will be tested within 48 hours of being loaded onto waste hauling trailers. The analytical results from each sampling round should be completed and reported within 5 working days of sampling. Until analysis is complete, the trailer cannot be used.

(4) *Define the scale of decision making* — The scale of decision making will be each container of waste ash.

(5) *Identify practical constraints on data collection* — The most important practical consideration that could interfere with the study is the ability to take samples from the fly ash that is stored in waste hauling trailers. Although the trailers have open access, special procedures and methods will have to be implemented for the samples to be representative of the entire depth of the ash. It has been suggested that core samples may be one practical solution to this problem. To get additional samples from each truck and to minimize the cost, compositing of core samples has been suggested.

Develop a Decision Rule — to define the parameter of interest, specify the action level and integrate previous DQO outputs into a single statement that describes a logical basis for choosing among alternative actions.

(1) *Specify the statistical parameter that characterizes the population of interest* — The planning team is interested in the true mean concentration of cadmium in the TCLP leachate for each container.

(2) *Specify the action level for the study* — The action level for the decision will be the RCRA regulatory standard for cadmium of 1.0 mg/L in the TCLP leachate.

(3) *Develop a decision rule (an "if...then..." statement)* — If the mean concentration of cadmium from the fly ash leachate in each container load is greater than 1.0 mg/L (using the TCLP method as defined in 40 CFR 261), then the waste will be considered hazardous and will be disposed of at a RCRA landfill. If the mean concentration of cadmium from the fly ash waste leachate is less than 1.0 mg/L (using the TCLP method as defined in 40 CFR 261), then the waste will be considered non-hazardous and will be disposed of in a sanitary landfill.

Specify Tolerable Limits on Decision Errors — the decision maker's tolerable decision error rates based on a consideration of the consequences of making a decision error.

(1) *Determine the possible range of the parameter of interest* — From analysis of records of similar studies of cadmium in environmental matrices, the range of the cadmium concentrations is expected to be from 0-2 mg/L. Therefore the mean concentration is expected to be between 0-2 mg/L for this investigation.

(2) *Identify the decision errors and choose the null hypothesis* —

(a) *Define both types of decision errors and establish the true state of nature for each decision error.* The planning team has determined that the two decision errors are (i) deciding that the waste is hazardous when it truly is not, and (ii) deciding that the waste is not hazardous when it truly is.

The true state of nature for decision error (i) is that the waste is not hazardous.

The true state of nature for decision error (ii) is that the waste is hazardous.

(b) *Specify and evaluate the potential consequences of each decision error.*

The consequences of deciding that the waste is hazardous when it truly is not will be that the incinerator company will have to pay more for the disposal of the fly ash at a RCRA facility than at a sanitary landfill.

Under the current Code of Federal Regulations, 40 CFR, Part 261, a solid waste can be considered "hazardous" if it meets specific criteria of ignitability, corrosivity, reactivity, and toxicity. One method that is used to determine if a solid substance, such as fly ash, meets the criteria for toxicity under the RCRA program regulations is to test a "representative sample" of the waste and perform a Toxicity Characteristic Leaching Procedure (TCLP) described in 40 CFR, Pt. 261, App. II. During this process, the solid fly ash will be "extracted" using an acid solution. The extraction liquid (the TCLP leachate) will then be subjected to tests for specific metals and compounds. For this example, the only concern is with the concentration of cadmium in the leachate. The primary benefit of the DQO Process will be to establish the data collection design needed to determine if the waste is hazardous under RCRA regulations within tolerable decision error rates.

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 between loads of ash leaving the facility. They have determined that each load is fairly homogeneous. There is a high variability between loads, however, due to the nature of the waste stream. Most of the fly ash produced is not hazardous and may be disposed of in a sanitary landfill. Thus, the company has decided that testing each individual waste load before it leaves the facility would be the most economical. Then they could send loads of ash that exceeded the regulated standards to the higher-cost RCRA landfills and continue to send the others to the sanitary landfill.

DQO Development

The following is a representative example of the output from each step of the DQO Process for the fly ash toxicity problem.

State the Problem — a description of the problem(s) and specifications of available resources and relevant deadlines for the study.

(1) *Identify the members of the planning team* — The members of the planning team will include the incineration plant manager, a plant engineer, a statistician, a quality assurance officer, an EPA representative who works within the RCRA program, and a chemist with sampling experience.

(2) *Identify the primary decision maker* — There will not be a primary decision maker; decisions will be made by consensus.

(3) *Develop a concise description of the problem* — The problem is to determine which loads should be sent to a RCRA landfill versus a sanitary landfill.

(4) *Specify available resources and relevant deadlines for the study* — While the project will not by constrained by cost, the waste generator (the incineration company) wishes to hold sampling costs below $2,500. They have also requested that the waste testing be completed within 1 week for each container load.

(4) *Assign probability values to points above and below the action level that reflect the tolerable probability for the occurrence of decision errors* — For this example, RCRA regulations allow a 5% decision error rate at the action level. The planning team has set the decision error rate to 5% from 1 mg/L to 1.5 mg/L and 1% from 1.5 mg/L to 2 mg/L as the consequences of health effects from the waste disposed of in the municipal landfill increase. On the other side of the action level, the planning team has set the tolerable probability of making a false negative error at 20% when the true parameter is from 0.25 to 0.75 mg/L and 10% when it is below 0.25 mg/L, based on both experience and an economic analysis that shows that these decision error rates are reasonable to balance the cost of sampling versus the consequence of sending clean ash to the RCRA facility (see Figure B-1).

Optimize the Design — select the most resource-effective data collection and analysis design for generating data that are expected to satisfy the DQOs. Optimizing the design is the one step of the DQO Process that will most likely be completed by a statistician or someone who has data collection design expertise. Using the case study as an example, the following section has been included to provide the reader with a background on the overall process that the statistician might follow to optimize the final data collection design.

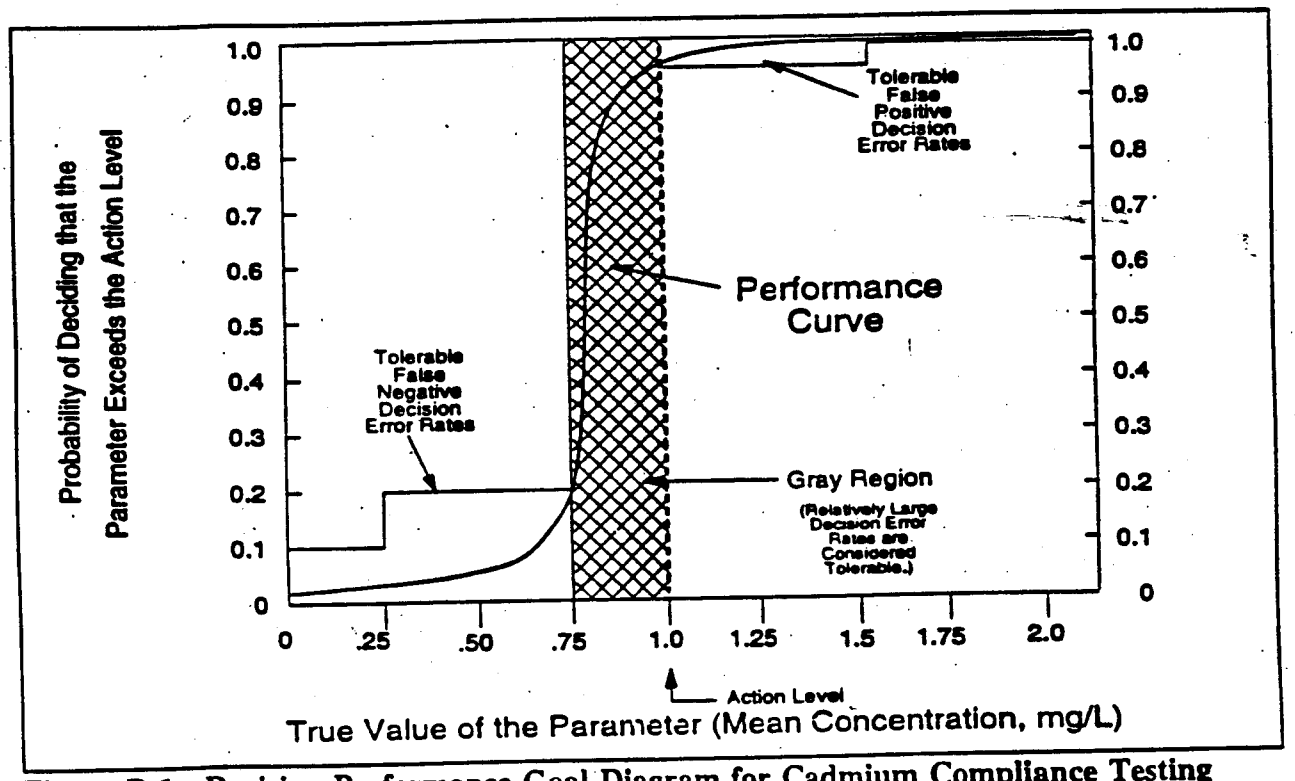


Figure B-1. Decision Performance Goal Diagram for Cadmium Compliance Testing Baseline Condition: Mean Exceeds Action Level.

(1) *Specify the characteristics that define the population of interest* — Fly ash waste from the hazardous waste incinerator will be analyzed. The fly ash should not be mixed with any other constituents except water that is used for dust control. Each load of ash should fill at least 70% of the waste trailer. In cases where the trailer is filled less than 70%, the trailer must wait on-site until more ash is produced and fills the trailer to the appropriate capacity.

(2) *Define the spatial boundary of the decision statement* —

(a) *Define the geographic area to which the decision statement applies.* Decisions will apply to each container load of fly ash waste.

(b) *When appropriate, divide the population into strata that have relatively homogeneous characteristics.* Stratification is not necessary since the waste ash is relatively homogeneous within each container.

(3) *Define the temporal boundary of the decision statement* —

(a) *Determine the timeframe to which the decision statement applies.* It will be assumed that the sampling data represent both the current and future concentration of cadmium within the ash.

(b) *Determine when to collect data.* Contained in the trucks, the waste does not pose a threat to humans or the environment. Additionally, since the fly ash is not subject to change, disintegration, or alteration, the decision about the waste characteristics does not warrant any temporal constraints. To expedite decision making, however, the planning team has placed deadlines on sampling and reporting. The fly ash waste will be tested within 48 hours of being loaded onto waste hauling trailers. The analytical results from each sampling round should be completed and reported within 5 working days of sampling. Until analysis is complete, the trailer cannot be used.

(4) *Define the scale of decision making* — The scale of decision making will be each container of waste ash.

(5) *Identify practical constraints on data collection* — The most important practical consideration that could interfere with the study is the ability to take samples from the fly ash that is stored in waste hauling trailers. Although the trailers have open access, special procedures and methods will have to be implemented for the samples to be representative of the entire depth of the ash. It has been suggested that core samples may be one practical solution to this problem. To get additional samples from each truck and to minimize the cost, compositing of core samples has been suggested.

In the case of the cadmium-contaminated ash, a fixed number of random grab samples would be selected and analyzed. Standard lab splits and QC samples would be taken according to standard procedures for the RCRA program. Each sample would be chosen randomly in three dimensions. A Student's t-test is suggested as a possible method for testing the statistical hypothesis.

(b) Composite Simple Random Sampling (composite sampling) — This type of sampling consists of taking multiple samples, physically combining (compositing) them, and drawing one or more subsamples for analysis. Composite samples are taken primarily when an average concentration is sought and there is no need to detect peak concentrations. By compositing the samples, researchers are able to sample a larger number of locations than if compositing was not used, while reducing the cost of analysis by combining several samples.

In the case of the cadmium-contaminated ash, a fixed number of random grab samples would be taken and composited. The number of grab samples contained in a composite sample (g) is also fixed. To determine sampling locations within the composite, a container would be divided into "g" equal-volume strata and samples would be chosen randomly within each strata. The use of strata ensure full coverage of each container. Standard lab splits and QC samples would be taken according to standard procedures for the RCRA program. A Student's t-test is suggested as the possible method for testing the statistical hypothesis.

(c) Sequential Sampling — Sequential sampling involves making several rounds of sampling and analysis. A statistical test is performed after each analysis to arrive at one of three possible decisions: reject the null hypothesis, accept the null hypothesis,[1] or collect more samples. This strategy is applicable when sampling and/or analysis costs are high, when information concerning sampling and/or measurement variability is lacking, when the waste and site characteristics of interest are stable over the timeframe of the sampling effort, and when the objective of the sampling is to test a single hypothesis. By taking samples in sequence, the researcher can hold down the cost of sampling and analysis.

In the case of the cadmium-contaminated ash, a sequential probability sample could be performed. The samples in each sampling round would be chosen randomly in three dimensions. If the decision to stop sampling has not been made before the number of samples required for the simple random sample are taken, sampling would stop at this point and the simple random sample test would be performed. Standard laboratory splits and QC samples would be taken according to standard procedures for the RCRA program. An approximate ratio test is

[1]Decide not to reject the null based on tolerable decision error limits.

The consequences of deciding that the waste is not hazardous when it truly is will be that the incinerator company will dispose of the waste in a sanitary landfill which could possibly endanger human health and the environment. In this situation, they may also be liable for future damages and environmental cleanup costs. Additionally, the reputation of the incinerator company may be compromised, jeopardizing its future profitability.

(c) *Establish which decision error has more severe consequences near the action level.* The planning team has concluded that decision error (ii) has the more severe consequences near the action level since the risk of jeopardizing human health outweighs the consequences of having to pay more for disposal.

(d) *Define the null hypothesis (baseline condition) and the alternative hypothesis and assign the terms "false positive" and "false negative" to the appropriate decision error.*

The baseline condition or null hypothesis (H_0) is "the waste is hazardous."

The alternative hypothesis (H_a) is "the waste is not hazardous."

The false positive decision error occurs when the null hypothesis is rejected when it is true. For this example, the false positive decision error occurs when the decision maker decides the waste is not hazardous when it truly is hazardous. The false negative decision error occurs when the null hypothesis is not rejected when it is false. For this example, the false negative decision error occurs when the decision maker decides that the waste is hazardous when it truly is not hazardous.

(3) *Specify a range of possible values of the parameter of interest where the consequences of decision errors are relatively minor (gray region)* — The gray region is the area adjacent to the action level where the planning team feels that the consequences of a false negative decision error are minimal. To decide how to set the width of the gray region, the planning team must decide where the consequences of a false negative decision error are minimal. Below the action level, even if the concentration of cadmium were very close to the action level, the monetary costs of disposing of the waste at a RCRA facility are the same as if the waste had a much lower concentration of cadmium. Clearly <u>any</u> false negative decision error (to the left of the action level) will cause the incinerator company and their customers to bear the cost of unnecessary expense (i.e., sending nonhazardous waste to a RCRA facility). The planning team, however, also realizes that they must define a reasonable gray region that balances the cost of sampling with risk to human health and the environment and the ability of measurement instruments to detect differences. Therefore the planning team has specified a width of 0.25 mg/L for this gray region based on their preferences to detect decision errors at a concentration of 0.75 mg/L (see Figure B-1).

- U.S. Environmental Protection Agency. 1994. *Statistical Methods for Evaluating the Attainment of Clean-up Standards: Volume 3: Reference-Based Standards for Soils and Solid Media.* EPA 230-R-94-004. Office of Policy, Planning and Evalutaion.

These formulas can also be found in many basic statistics textbooks. Different formulas are necessary for each data collection design, for each parameter, and for each statistical test. These formulas are generally a function of α; β; the detection difference, Δ (delta); and the standard deviation, σ. The detection difference, Δ, is defined to be the difference between the action level (AL) and the other bound of the gray region (U); i.e., Δ = AL - U. In this case the standard deviation was derived from pilot data under approximately the same conditions as expected for the real facility.

For example, a formula for computing the sample size necessary to meet the DQO constraints for comparing a mean against a regulatory threshold, when a simple random sample is selected, is:

$$n = \frac{\sigma^2(z_{1-\beta}+z_{1-\alpha})^2}{\Delta^2}+(0.5)z_{1-\alpha}^2$$

where:

σ^2 = estimated variance in measurements (from pilot study)
n = number of samples required,
z_p = the pth percentile of the standard normal distribution (from standard statistical tables), and
Δ = U - AL

Simple Random Sample — Using the formula above, it was determined that 37 samples are necessary to achieve the specified limits on decision errors. This sampling plan satisfies all the DQOs including budget, schedule, and practical constraints.

Composite Sampling — To determine sample sizes for a composite sample, it is necessary to compute the number of composites samples, n; the number of samples, g, within each composite; and the number of subsamples, m, to be measured for each composite. Usually m=1; however, since this design is to be used repeatedly, it is suggested that two subsamples from each composite sample be measured to estimate composite variability, which can then be used to re-optimize the number of samples m and g.

For a composite sample, with random sample locations, it has been determined that eight composite samples of eight samples each are sufficient to meet the limits on decision errors that have been specified. This design is more than sufficient to

Overview

Developing a data collection design requires an understanding of the sampled medium and the information that was generated in previous DQO steps. The statistician's job is to review the background information, determine the appropriate statistical application to adequately solve the problem, and develop one or more appropriate data collection designs. Once this is complete, the statistician will compare the cost and performance of the different data collection designs. This process can be broken down into five distinct steps:

(1) Review the DQO outputs and existing environmental data.

(2) Develop general data collection design alternatives.

(3) For each data collection design alternative, select the optimal sample size that satisfies the DQOs.

(4) Select the most resource-effective data collection design that satisfies all of the DQOs.

(5) Document the operational details and theoretical assumptions of the selected design in the sampling and analysis plan.

Activities

(1) *Review the DQO outputs and existing environmental data* — Because the statistician has participated in the DQO Process for this problem, there is no need to review the DQO outputs further. The only existing data relevant to this problem are the pilot study data. Based on the pilot study, the incineration company has determined that each load of ash is fairly homogeneous, and has estimated the standard deviation in the concentration of cadmium within loads of ash to be 0.6 mg/L.

(2) *Develop general data collection design alternatives* — Generally, the design alternatives are based on a combination of design objectives developed in previous DQO Process steps and knowledge of statistical parameters about the medium or contaminant. Below are four examples of possible designs that could apply to the case study:

(a) Simple Random Sampling — The simplest type of probability sample is the simple random sample. With this type of sampling, every possible point in the sampling medium has an equal chance of being selected. Simple random samples are used primarily when the variability of the medium is relatively small and the cost of analysis is relatively inexpensive. Simple random sample locations are generally developed through the use of a random number table or through computer generation of pseudo-random numbers.

This design meets all of the required limits on decision errors. The cost of this design is based on the cost of selecting ($10) and analyzing ($150) a sample. Eight samples will be used to make each composite sample for a sampling cost of $80; two subsamples will be analyzed from this composite sample for a cost of $300. Therefore, each composite sample will cost $380. The total cost of this design is:

$$Cost_{CS} = 8 \times \$380 = \$3040.$$

Sequential Sampling — Sequential sampling will be performed similarly to random sampling. The primary difference is that the ultimate number of samples will be determined by the results of one or more sampling rounds.

This design has the potential to reduce the number of samples required in the simple random sampling design and still meet the decision error limits. The average costs of the two decisions are used below:

The ash is hazardous:	16 × ($160) = $2,560
The ash is non-hazardous:	22 × ($160) = $3,520

To determine the expected cost, estimate the number of loads of ash that should be sent to a RCRA facility versus the number of loads that can be sent to a municipal facility. Suppose 25% of the loads are hazardous and should be sent to a RCRA facility. Then the expected cost (EC_{SS}) of this design should be

EC_{SS} = 0.25 × (cost of sampling when ash is hazardous) + (0.75 × cost of sampling when ash is non-hazardous)

$$= 0.25 \times (\$2{,}560) + 0.75 \times (\$3{,}520) = \$\ 3{,}280$$

Selection of a Design

Because the simple random sampling design requires that many samples be taken and analyzed, it is inefficient for the goals of this study. Sampling will cost almost as much to determine whether the waste is hazardous or nonhazardous as it would cost to send all the waste to a RCRA hazardous waste landfill. Therefore, this decision is not resource-effective.

The sequential data collection design is more resource-effective than the simple random sampling design. The potential savings over sending all waste to a RCRA hazardous waste facility is $6,750 - $3,280 = $3,470. The site owner has expressed disapproval for this sampling plan because of the time it may take before a decision can be made. If the ash was not homogeneous within a container, however, this data collection design may be the design of choice.

suggested after each round of sampling is complete to decide whether or not to conclude that the waste is hazardous or to continue sampling.

(d) Stratified Random Sampling — Stratified sampling involves dividing the study area into two or more non-overlapping subsets (strata) which cover the entire volume to be sampled. These strata should be defined so that physical samples within a stratum are more similar to each other than to samples from other strata. Sampling depth, concentration level, previous cleanup attempts, and confounding contaminants can be used as the basis for creating strata. Once the strata have been defined, each stratum is then sampled separately using one of the above designs. Stratification is often used to ensure that important areas of a site are represented in the sample. In addition, a stratified random sample may provide more precise estimates of contaminant levels than those obtained from a simple random sample. Even with imperfect information, a stratified sample can be more resource-effective.

Since the incineration company has already determined that each load of ash is fairly homogeneous, stratification does not have any advantages over a simple random sample. In addition, since the company has decided to test each waste load individually before it leaves the facility, stratifying each waste load would be difficult and unnecessary. Therefore, this data collection design will not be considered further.

(3) *For each data collection design alternative, select the optimal sample size that satisfies the DQOs* — The formula for determining the sample size (number of samples to be collected) is chosen based on the hypothesis test and data collection design. Standard formulas can be found in several references, including:

- Cochran, W. 1977. *Sampling Techniques.* New York: John Wiley.
- Desu, M.M., and D. Raghavarao. 1990. *Sample Size Methodology.* San Diego, CA: Academic Press.
- Gilbert, Richard O. 1987. *Statistical Methods for Environmental Pollution Monitoring.* New York: Van Nostrand Reinhold.
- U.S. Environmental Protection Agency. 1989. *Methods for Evaluating the Attainment of Cleanup Standards: Volume 1: Soils and Solid Media.* EPA 230/02-89-042, Office of Policy, Planning and Evaluation.
- U.S. Environmental Protection Agency. 1992. *Methods for Evaluating the Attainment of Cleanup Standards: Volume 2: Ground Water.* EPA 230-R-92-014, Office of Policy, Planning and Evaluation.

APPENDIX C

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 of Appendix B.

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

$$Z = \frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} \tag{1}$$

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

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

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

Note that

$$T = \frac{[(\bar{X}-\mu)+(\mu-AL)]\sqrt{n}}{\sigma} = Z+\varepsilon(\mu) \tag{3}$$

achieve the specified limits on decision errors and satisfies all the DQOs including budget, schedule, and practical constraints.

Sequential Sampling — For the purposes of comparing costs, the average number of samples in a sequential sampling design can be estimated, but these estimates are only averages. The average sample size for concluding that the waste is hazardous is 16 and the average sample size for concluding the waste is not hazardous is 22. The average sizes are different because the burden of proof is placed on disproving the null hypothesis, thus, more samples on average are required to prove that the alternative hypothesis (the waste is not hazardous) is true. However, these sample sizes are only averages. In some cases, fewer samples are necessary; in others, more may be necessary. This sampling plan satisfies all the DQOs including budget, schedule, and practical constraints.

(4) *Select the most resource-effective data collection design that satisfies the DQOs* — Compare the overall efficiency of each model and choose the one that will solve the problem most effectively.

Cost Estimates for Each Design

First, the costs for the three designs alternatives will be evaluated:

Simple Random Sampling — A simple random sampling scheme can be implemented for each load of fly ash by first generating three-dimensional random sampling points. This can most easily be done by using a computer. Samples can then be taken using a special grab sampler which will be forced into the ash, opened to take the sample, then closed and removed. The difficulty with this type of sampling scheme is measuring sampling locations in three dimensions, and it may be difficult to gain access to the correct sampling locations.

This design meets all of the required limits on decision errors. The cost of this design is calculated based on the assumed cost of selecting a sample ($10), and the cost of analyzing a sample ($150). Since 37 samples need to be taken and analyzed, the cost of this design is:

$$\begin{aligned} Cost_{SRS} &= 37 \times \$10 + 37 \times \$150 \\ &= \$370 + \$5550 = \$5920 \end{aligned}$$

Composite Sampling — Composite sampling will be performed similarly to simple random sampling except that after eight random samples are collected (one from each stratum), they will be combined and homogenized. Two sample aliquots for analysis will then be drawn from the homogenized mixture. This process will be repeated eight times.

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:

$$n = \frac{(z_{1-\alpha}+z_{1-\beta})^2\sigma^2}{\Delta^2} + \frac{1}{2}z_{1-\alpha}^2. \tag{8}$$

In practice, since σ is unknown, a prior estimate of it must be used in (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

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

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

$$Pr\left[Z>k\sqrt{n}/\sqrt{1+k^2/2}\right] = \alpha, \tag{10}$$

$$i.e.,\ z_{1-\alpha} = k\sqrt{n}/\sqrt{1+k^2/2}. \tag{11}$$

One can show that (11) is equivalent to

$$1/[1+k^2/2] = 1-z_{1-\alpha}^2/2n. \tag{12}$$

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 $\leq$ 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

$$\frac{\bar{X}-kS-(U-k\sigma)}{\left(\sigma/\sqrt{n}\right)\sqrt{1+k^2/2}} \leq \frac{AL-(U-k\sigma)}{\left(\sigma/\sqrt{n}\right)\sqrt{1+k^2/2}}. \tag{13}$$

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

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

which implies

$$-z_{1-\beta} = z_\beta = \frac{(AL-U)+k\sigma}{\left(\sigma/\sqrt{n}\right)\sqrt{1+k^2/2}}. \tag{14}$$

Subtracting (14) from (11) yields

$$z_{1-\alpha}+z_{1-\beta} = \frac{(U-AL)}{\left(\sigma/\sqrt{n}\right)\sqrt{1+k^2/2}}, \tag{15}$$

or

$$\frac{(z_{1-\alpha}+z_{1-\beta})\sigma}{(U-AL)} = \frac{\sqrt{n}}{\sqrt{1+k^2/2}}. \tag{16}$$

Substituting (12) into the denominator on the right side of (16) yields

$$\frac{(z_{1-\alpha}+z_{1-\beta})\sigma}{(U-AL)} = \sqrt{n}\sqrt{1-z_{1-\alpha}^2/2n}. \tag{17}$$

Squaring both sides of (17) and solving for n yields equation (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.* Vol. 35, No. 4.

APPENDIX D

GLOSSARY OF TERMS

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

alternative hypothesis: *See hypothesis.*

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

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

data collection design: A data collection design specifies the configuration of the environmental monitoring effort to satisfy the DQOs. It includes the types of samples or monitoring information to be collected; where, when, and under what conditions they should be collected; what variables are to be measured; and the Quality Assurance and Quality Control (QA/QC) components that ensure acceptable sampling design error and measurement error to meet the decision error rates specified in the DQOs. The data collection design is the principal part of the QAPP.

Data Quality Assessment (DQA) Process: a statistical and scientific evaluation of the data set to assess the validity and performance of the data collection design and statistical test, and to establish whether a data set is adequate 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 the tolerable levels of potential decision errors that will be used as the basis for establishing the quality and quantity of data needed to support decisions.

Data Quality Objectives Process: a Quality Management tool based on the Scientific Method, developed by the U.S. Environmental Protection Agency to facilitate the planning of environmental data collection activities. The DQO Process enables planners to focus their planning efforts by specifying the intended use of the data (the decision), the decision criteria (action level), and the decision maker's tolerable decision error rates. The products of the DQO Process are the DQOs.

decision error: an error made when drawing an inference from data in the context of hypothesis testing, such that variability or bias in the data mislead the decision maker to draw a conclusion that is inconsistent with the true or actual state of the population under study. *See also false negative decision error, false positive decision error.*

defensible: the ability to withstand any reasonable challenge related to the veracity, integrity, or quality of the logical, technical, or scientific approach taken in a decision making process.

false negative decision error: a false negative decision error occurs when the decision maker does not reject the null hypothesis when the null hypothesis actually is false. In statistical terminology, a false negative decision error is also called a Type II error. The measure of the size of the error is expressed as a probability, usually referred to as "beta (β)"; this probability is also called the complement of power.

false positive decision error: a false positive decision error occurs when a decision maker rejects the null hypothesis when the null hypothesis actually is true. In statistical terminology, a false positive decision error is also called a Type I error. The measure of the size of the error is expressed as a probability, usually referred to as "alpha (α)," the "level of significance," or "size of the critical region."

gray region: a range of values of the population parameter of interest (such as mean contaminant concentration) where the consequences of making a decision error are relatively minor. The gray region is bounded on one side by the action level.

hypothesis: a tentative assumption made to draw out and test its logical or empirical consequences. In hypothesis testing, the hypothesis is labeled "null" or "alternative", depending on the decision maker's concerns for making a decision error.

limits on decision errors: the tolerable decision error probabilities established by the decision maker. Potential economic, health, ecological, political, and social consequences of decision errors should be considered when setting the limits.

mean: (i) a measure of central tendency of the population (population mean), or (ii) the arithmetic average of a set of values (sample mean).

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

median: 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. The median is the 50th percentile.

medium: a substance (e.g., air, water, soil) which 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 hypothesis.*

parameter: a numerical descriptive measure of a population.

percentile: the specific value of a distribution that divides the distribution such that p percent of the distribution is equal to or below that value. Example for p=95: "The 95th percentile is X" means that 95% of the values in the population (or statistical sample) are less than or equal to X.

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

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

power function: the probability of rejecting the null hypothesis (H_o) over the range of possible population parameter values. The power function is used to assess the goodness of a hypothesis test or to compare two competing tests.

quality assurance (QA): an integrated system of management activities involving planning, quality control, quality assessment, reporting, and quality improvement to ensure that a product or service (e.g., environmental data) meets defined standards of quality with a stated level of confidence.

Quality Assurance Project Plan (QAPP): a formal technical document containing the detailed QA, QC and other technical procedures for assuring the quality of environmental data prepared for each EPA environmental data collection activity and approved prior to collecting the data.

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

Quality Management Plan (QMP): a formal document describing the management policies, objectives, principles, organizational authority, responsibilities, accountability, and implementation protocols of an agency, organization, or laboratory for ensuring quality in its products and utility to its users. In EPA, QMPs are submitted to the Quality Assurance Management Staff (QAMS) for approval.

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

[1]sample: a single item or specimen from a larger whole or group, such as any single sample of any medium (air, water, soil, etc.).

[2]sample: a set of individual samples (specimens or readings), drawn from a population, whose properties are studied to gain information about the whole.

sampling: the process of obtaining representative samples and/or measurements of a subset of a population.

sampling design error: the error due to observing only a limited number of the total possible values that make up the population being studied. It should be distinguished from errors due to imperfect selection; bias in response; and errors of observation, measurement, or recording, etc.

scientific method: the principles and processes regarded as necessary for scientific investigation, including rules for concept or hypothesis formulation, conduct of experiments, and validation of hypotheses by analysis of observations.

standard deviation: the square root of the variance.

statistic: a function of the sample measurements; e.g., the sample mean or standard deviation.

statistical test: any statistical method that is used to determine which of several hypotheses are true.

total study error: the combination of sampling design error and measurement error.

true: being in accord with the actual state of affairs.

Type I error: A Type I error occurs when a decision maker rejects the null hypothesis when it is actually true. *See false positive decision error.*

Type II error: A Type II error occurs when the decision maker fails to reject the null hypothesis when it is actually false. *See false negative decision error.*

variable: The attribute of the environment that is indeterminant.

variance: a measure of (i) the variability or dispersion in a population (population variance), or (ii) the sum of the squared deviations of the measurements about their mean divided by the degrees of freedom (sample variance).