APPENDIX B
THE DATA QUALITY OBJECTIVES PROCESS

B.1 Introduction

This appendix provides information about the basic framework of the DQO process (ASTM 5792; EPA, 2000; NRC, 1998; MARSSIM, 2000). The DQO planning process empowers both data users and data suppliers to take control and resolve issues in a stepwise fashion. It brings together at the right time all key players from the data user and data supplier constituencies and enables each participant to play a constructive role in clearly defining:

- The problem that requires resolution;
- What type, quantity, and quality of data the decisionmaker needs to resolve that problem;
- Why the decisionmaker needs that type and quality of data;
- How much risk of making a wrong decision is acceptable; and
- How the decisionmaker will use the data to make a defensible decision.

The DQO process provides a logic for setting well-defined, achievable objectives and developing a cost-effective, technically sound sampling and analysis design. It balances the data user's tolerance for uncertainty with the available resources for obtaining data. The number of visible and successful applications of the DQO process has proven its value to the environmental community. The DQO process is adaptable depending on the complexity of the project and the input from the decisionmakers. Some users have combined DQO planning with remedy selection for restoration projects (e.g., DOE's Streamlined Approach for Environmental Restoration—see Section A.5 in Appendix A). Other users have integrated their project scoping meetings with the DQO process. Much of the information that is developed during the DQO process is useful for developing the project plan documents (Chapter 4) and implementing the data validation process (Chapter 8) and the data quality assessment (DQA) process (Chapter 9).

Since its inception, the term "data quality objectives" has been adopted by many organizations, and the definition has been adapted and modified (see box on next page). Throughout this document, MARLAP uses EPA's (2000) definition of 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."

Contents

Definitions of Data Quality Objectives

(1) Statements on the level of uncertainty that a decisionmaker is willing to accept in the results derived from environmental data (ASTM 5283; EPA, 1986).
(2) 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 (EPA, 2000).
(3) Qualitative and quantitative statements derived from the DQO process describing the decision rules and the uncertainties of the decision(s) within the context of the problem(s) (ASTM D5792).
(4) Qualitative and quantitative statements that specify the quality of the data required to support decisions for any process requiring radiochemical analysis (radioassay) (ANSI N42.23).

B.2 Overview of the DQO Process

The DQO process (Figure B.1) consists of seven steps (EPA, 2000). In general, the first four steps require the project planning team to define the problem and qualitatively determine required data quality. The next three steps establish quantitative performance measures for the decision and the data. The final step of the process involves developing the data collection design based on the DQOs, which is dependent on a clear understanding of the first six steps.

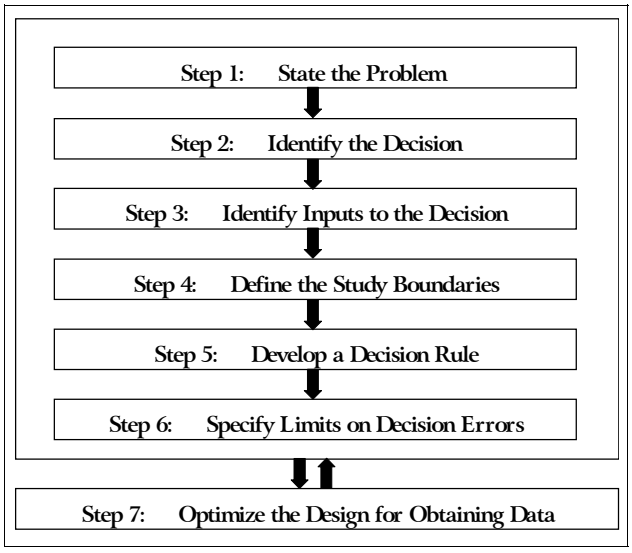

FIGURE B.1 — Seven steps of the DQO process

Although the DQO process is described as a sequence of steps, it is inherently iterative. The output from each step influences the choices that will be made in subsequent steps. For instance, a decision rule cannot be created without first knowing the problem and desired decision. Similarly, optimization of the sampling and analysis design generally cannot occur unless it is clear what is being optimized —the results of the preceding steps. Often the outputs of one step will trigger the need to rethink or address issues that were not evaluated thoroughly in prior steps. These iterations lead to a more focused sampling and analysis design for resolving the defined problem. The first six steps should be completed before the sampling and analysis design is developed, and every step should be completed before data collection begins. The DQO process is considered complete with the approval of an optimal design for sampling and analysis to support a decision or when available historical data are sufficient to support a decision.

In practice, project planning teams often do a cursory job on the first four steps, wanting to get into technical design issues immediately. Without carefully defining the problem and the desired result, the project planning team may develop a design that is technically sound but answers the wrong question, or answers the questions only after the collection of significant quantities of unnecessary data. Time spent on the first four steps is well spent. Extra effort must be given to assure that Steps 1 to 4 are adequately addressed.

When applying the DQO process, or any planning approach, it is important to document the outputs of each step to assure that all participants understand and approve the interim products, and that they have a clear record of their progress. It is sometimes useful to circulate an approval copy with signature page to ensure agreement of the stakeholders.

B.3 The Seven Steps of the DQO Process

Each step of the DQO process will be discussed in the following sections. Not all items will be applicable to every project. The project planning team should apply the concepts that are appropriate to the problem.

B.3.1 DQO Process Step 1: State the Problem

The first step is to define the problem clearly. The members of the project planning team present their concerns, identify regulatory issues and threshold levels, and review the site history. The project planning team should develop a concise description of the problem. Some elements to include in the description might be the study objectives, regulatory context, groups who have an interest in the study, funding and other resources available, previous study results, and any obvious sampling design constraints. The more facts, perceptions and concerns of the key stakeholders—including important social, economic, or political issues—that are identified during this step, the better the chances are that the issues driving the decisions and actions will be identified.

The primary decisionmaker should be identified. The resources and relevant deadlines to address the problem are also defined at this time. If possible, a "project conceptual model" should be developed. This will help structure and package the diverse facts into an understandable picture of what the various issues are and how those issues can be focused into a specific problem. The expected outputs of Step 1 are:

- A conceptual model that packages all the existing information into an understandable picture of the problem;

- A list of the project planning team members and identification of the decisionmaker;

- A concise description of the problem; and

- A summary of available resources and relevant deadlines for the study.

B.3.2 DQO Process Step 2: Identify the Decision

During Step 2 of the DQO process, the project planning team defines what decision must be made or what question the project will attempt to resolve. The decision (or question) could be simple, like whether a particular discharge is or is not in compliance, or the decision could be complex, such as determining if observed adverse health is being caused by a nonpoint source discharge. Linking the problem and the decision focuses the project planning team on seeking only that information essential for decisionmaking, saving valuable time and money.

The result may be a comprehensive decision for a straightforward problem, or a sequence of decisions for a complex problem. For complex problems with multiple concerns, these concerns should be ranked in order of importance. Often a complex concern is associated with a series of decisions that need to be made. Once these decisions have been identified, they should be sequenced in a logical order so the answer to one decision provides input in answering the next decision. It may be helpful to develop a logic-flow diagram (decision framework), arraying each element of the issue in its proper sequence along with its associated decision that requires an answer.

The term "action level" is used in this document to denote the numerical value that will cause the decisionmaker to choose one of the alternative actions. The action level may be a derived concentration guideline level, background level, release criteria, regulatory decision limit, etc. The action level is often associated with the type of media, analyte, and concentration limit. Some action levels, such as release criteria for license termination, are expressed in terms of dose or risk. The release criteria typically are based on the total effective dose equivalent (TEDE), the committed effective dose equivalent (CEDE), risk of cancer incidence (morbidity), or risk of cancer death (mortality), and generally cannot be measured directly. A radionuclide-specific predicted concentration or surface area concentration of specific nuclides that can result in a dose (TEDE or CEDE) or specific risk equal to the release criterion is called the "derived concentration guideline level" (DCGL). A direct comparison can be made between the project's analytical measurements and the DCGL (MARSSIM, 2000).

The project planning team should define the possible actions that may be taken to solve the problem. Consideration should be given to the option of taking no action. A decision statement can then be developed by combining the decisions and the alternative actions. The decision rule and the related hypothesis test will be more fully developed in the DQO process at Steps 5 and 6.

By defining the problem and its associated decision clearly, the project planning team has also begun to define the inputs and boundaries (DQO process Steps 3 and 4). At the end of Step 2, the

project planning team has:

- Identified the principal decisions or questions;
- Defined alternative actions that could be taken to solve the problem based on possible answers to the principal decisions and questions;
- Combined the principal decisions and questions and the alternative actions into decision statements that expresses a choice among alternative actions; and
- Organized multiple decisions.

B.3.3 DQO Process Step 3: Identify Inputs to the Decision

During Step 3, the project planning team makes a formal list of the specific information required for decisionmaking. The project planning team should determine what information is needed and how it can be acquired. The project planning team should specify if new measurements are required for the listed data requirements. The data required are based on outcomes of discussion during the previous two steps. The project planning team should define the basis for setting the action level. Depending on the level of detail of the discussion during the previous steps, then efforts associated with Step 3 may be primarily to capture that information. If the first two steps have not defined the inputs with enough specificity, then those inputs should be defined here. However, before going further, the output should be reviewed to assure that the problem, the decision steps and the input are compatible in complete agreement.

An important activity during Step 3 is to determine if the existing data or information, when compared with the desired information, has significant gaps. If no gaps exist, then the existing data or information may be sufficient to resolve the problem and make the decision. (Although there may be no gaps in the data, the data may not have enough statistical power to resolve the action level. See Step 6 for more discussion.) In order to optimize the use of resources, the project planning team should maximize the use of historical information. If new data are required, then this step establishes what new data (inputs) are needed. The specific environmental variable or characteristic to be measured should be identified. The DQO process clearly links sampling and analysis efforts to an action and a decision. This linkage allows the project planning team to determine when enough data have been collected.

If the project planning team determines that collection of additional data is needed, the analytical laboratory acquisition strategy options should be considered at this stage. Identifying suitable contracting options should be based on the scope, schedule, and budget of the project, and the capability and availability of laboratory resources during the life of the project, and other technical considerations of the project. If an ongoing contract with a laboratory is in place, it is advisable to involve them with the radioanalytical specialists as early as possible.

The project planning team should ensure that there are analytical protocols available to provide acceptable measurements. If analytical methods do not exist, the project planning team will need to consider the resources needed to develop a new method, reconsider the approach for providing input data, or perhaps reformulate the decision statement.

The expected outputs of Step 3 are:

- A list of information needed for decisionmaking;
- Determination of whether data exist and are sufficient to resolve the problem;
- Determination of what new data, if any, are required;
- Definition of the characteristics that define the population and domain of interest;
- Definition of the basis for the action level;
- Confirmation that appropriate analytical protocols exist to provide the necessary data; and
- A review of the planning output to assure the problem, decision, and inputs are fully linked.

B.3.4 DQO Process Step 4: Define the Study Boundaries

In Step 4, the project planning team specifies the spatial and temporal boundaries covered by the decision statement. The spatial boundaries define the physical aspects to be studied in terms of geographic area, media, and any appropriate subpopulations (e.g., an entire plant, entire river basin, one discharge, metropolitan air, emissions from a power plant). When appropriate, divide the population into strata that have relatively homogeneous characteristics. The temporal boundaries describe the time frame the study data will represent (e.g., possible exposure to local residents over a 30-year period) and when samples should be taken (e.g., instantaneous samples, hourly samples, annual average based on monthly samples, samples after rain events). Changing conditions that could impact the success of sampling and analysis and interpretation need to be considered. These factors include weather, temperature, humidity, or amount of sunlight and wind.

The scale of the decision is also defined during this step. The selected scale should be the smallest, most appropriate subset of the population for which decisions will be made based on the spatial or temporal boundaries. During Step 4, the project planning team also should identify practical constraints on sampling and analysis that could interfere with full implementation of the data collection design. These include time, personnel, equipment, and seasonal or meteorological conditions when sampling is not possible or may bias the data.

In practice, the study boundaries are discussed when the project planning team and decision-maker agree on the problem and its associated decision. For instance, a land area that may be contaminated or a collection of waste containers would be identified as part of the problem and decision definition in Steps 1 and 2. The boundaries also would be considered when determining inputs to the decision in Step 3. If the study boundaries had not been addressed before Step 4 or if new issues were raised during Step 4, then Steps 1, 2, and 3 should be revisited to determine

how Step 4 results are now influencing the three previous steps.

The outputs of Step 4 are:

- A detailed description of the spatial and temporal boundaries of the problem; and
- Any practical constraints that may interfere with the sampling and analysis activities.

B.3.5 Outputs of DQO Process Steps 1 through 4 Lead Into Steps 5 through 7

At this stage in the DQO process, the project planning team has defined with a substantial degree of detail the problem, its associated decision, and the inputs and boundaries for addressing that problem. The project planning team knows whether it needs new data to fill specific gaps and what that data should be. The remaining three steps are highly technical and lead to the selection of the sampling and analysis design. Even when new data are not required (i.e., a data collection design is not needed), the project planning team should continue with Steps 5 and 6 of the DQO process. By establishing the formal decision rule and the quantitative estimates of tolerable decision error rates, the project planning team is assured that consensus has been reached on the actions to be taken and information to establish criteria for the DQA process.

It is important to emphasize that every effort must be made to assure that Steps 1 through 4 are adequately addressed. If the necessary time is taken in addressing the first four steps carefully and assuring consensus among the project planning team, then the three remaining steps are less difficult.

B.3.6 DQO Process Step 5: Develop a Decision Rule

In Step 5, the project planning team determines the appropriate statistical parameter that characterizes the population, specifies the action level, and integrates previous DQO process outputs into a single "if ..., then ..." statement (called a "decision rule") that describes a logical basis for choosing among alternative actions.

The four main elements to the decision rule are:

A. THE PARAMETER OF INTEREST. A descriptive measure (e.g., mean, median, or proportion) that specifies the characteristic or attribute that the decisionmaker would like to know and that the data will estimate. The characteristics that define the population and domain of interest was established in Step 3.

B. THE SCALE OF DECISIONMAKING. The smallest, most appropriate subset for which decisions will be made. The scale of decisionmaking was defined in Step 4.

C. THE ACTION LEVEL. A threshold value of the parameter of interest that provides the criterion

for choosing among alternatives. Action levels may be based on regulatory standards or they may be derived from project- and analyte-specific criteria such as dose or risk analysis. The basis for the action level was determined in Step 3.

D. THE ALTERNATIVE ACTIONS. The actions the decisionmaker would take, depending on the "true value" of the parameter of interest. The alternative actions were determined in Step 2.

The decision rule is a logical, sequential set of steps to be taken to resolve the problem. For example, "If one or more conditions exists then take action 1, otherwise take action 2."

The outputs of Step 5 are:

- The action level;
- The statistical parameter of interest; and
- An "if ..., then ..." statement that defines the conditions that would cause the decisionmaker to choose among alternative courses of action.

PROCEDURE FOR DEVELOPING A DECISION RULE

The outcome of a decision rule is a result: often to take action or not to take action. The decision rule is an "If..., then..." statement that defines the conditions that would cause the decisionmaker to choose an action. The decision rule establishes the exact criteria for making that choice. There are four main elements to a decision rule:

A. The *parameter of interest*. For example, the mean or median of the concentration of an analyte.
B. The *area over which the measurements are taken*. For example, in MARSSIM, a survey unit.
C. The *action level*. For example, in MARSSIM, the action level is called the DCGL.
D. *Alternative actions*. For example, if the mean is greater than the action level, then corrective action must be taken, otherwise the survey unit may be released.

A decision rule is action oriented, so a decision rule has the general form:

If the value of parameter A, over the area B, is greater than C, then take action D, otherwise take action D*.

For example, if:

(A) the true *mean concentration of* ^{238}U in the
(B) *surface soil* of the survey unit is greater than

(C) *30 pCi/g*, then
(D) *remove the soil* from the site; otherwise,
(D*) *leave the soil* in place.

The decisionmaker and planning team should be comfortable with the decision rule regarding the criteria for taking action before any measurements are taken. The input to a decision rule is the result of measurements. A decision will be made, and action taken, based upon those results.

There is uncertainty with every scientific measurement taken. Sampling uncertainty is due to the natural spatial and temporal variation in contaminant concentrations across a site. Measurement uncertainty is the variability in a combination of factors that arise during sample analysis. Because there is uncertainty in measurement results, the decision based on them could be incorrect. Controlling decision error is the subject of Step 6 of the DQO process.

B.3.7 DQO Process Step 6: Specify the Limits on Decision Errors

In this step, the project planning team assesses the potential consequences of making a wrong decision and establishes a tolerable level for making a decision error. The project planning team defines the types of decision errors (Type I and II) and the tolerable limits on the decision error rates. In general, a Type I error is deciding against the default assumption (the null hypothesis) when it is actually true; a Type II error is not deciding against the null hypothesis when it is actually false (see Attachment B1 and Appendix C for detailed discussions). The limits imposed on the probability of making decision errors will be used to establish measurement performance criteria for the data collection design.

Traditionally, the principles of statistical hypothesis testing have been used to determine tolerable levels of decision error rates. Other approaches applying decision theory have been applied (Bottrell et al., 1996a, b). Based on an understanding of the possible consequences of making a wrong decision in taking alternative actions, the project planning team chooses the null hypotheses and judges what decision error rates are tolerable for making a Type I or Type II decision error.

The project planning team also specifies a range of possible values where the consequences of decision errors are relatively minor (the gray region). Specifying a gray region is necessary because variability in the population and imprecision in the measurement system combine to produce variability in the data such that the decision may be "too close to call" when the true value is very near the action level. The width of the gray region establishes the distance from the action level where it is most important that the project planning team control Type II errors. For additional information on the gray region, hypothesis testing, and decision errors, see EPA (2000) and NRC (1998).

The tolerable decision error rates are used to establish performance goals for the data collection

design. Overall variability in the result can be attributed to several sources, including sample location, collection, and handling; laboratory handling and analysis; and data handling and analysis. In many environmental cases, sampling is a much larger source of uncertainty than laboratory analyses. The goal is to develop a sampling and analysis design that reduces the chance of making a wrong decision. The greater certainty demanded by the decisionmakers, the more comprehensive and expensive the data collection process is likely to be. In this step, the project planning team has to come to an agreement on how to determine acceptable analytical uncertainty and how good the overall data results are required to be. The team has to reach a consensus on the trade off between the cost of more information and the increased certainty in the resulting decision.

Often the project planning team does not feel comfortable with the concepts and terminology of hypothesis testing (Type I and Type II errors, gray region, critical region, tolerable decision error rates). As a result, the project planning team may have difficulty with (or want to skip) this step of the directed planning process. If these steps are skipped or insufficiently addressed, it is more likely that the data will not be of the quality needed for the project. Attachment B1 gives additional guidance on these concepts. MARLAP recommends that for each radionuclide of concern, an action level, gray region, and limits on decision error rates be established during a directed planning process. A stepwise procedure for accomplishing this is given at the end of this section.

Figure B.2 summarizes the outputs of the decisions made by the project planning team in a decision performance goal diagram (EPA, 2000). The horizontal axis represents the (unknown) true value of the parameter being estimated. The vertical axis represents the decisionmaker's desired probability of concluding that the parameter exceeds an action limit. The "gray region" (bounded on one side by the action level) defines an area where the consequences of decision error are relatively minor (in other words, it defines how big a divergence from the action level we wish to distinguish). The gray region is related to the desired precision of the measurements. The height of the indicated straight lines to the right and left of the gray region depict the decisionmaker's tolerance for Type I and Type II errors.

For purposes of this example, the default assumption (null hypothesis) was established as "the measured concentration exceeds the action level" (Figure B.2a). A Type I error consists in making a decision *not* to take action (e.g., remediate) when that action was in fact required (e.g., analyte concentrations are really above an action level). The desired limit on the probability of making a Type I error is set at 5 percent if the true concentration is between 100 and 150 and at 1 percent if the true concentration exceeds 150. A Type II error is understood as taking an action when in fact that action is not required (e.g., analyte concentrations are really below the action level). The desired limit on the probability of making a Type II error is set at 5 percent if the true concentrations is less than 25 and 10 percent if the true concentrations is between 25 and 75.

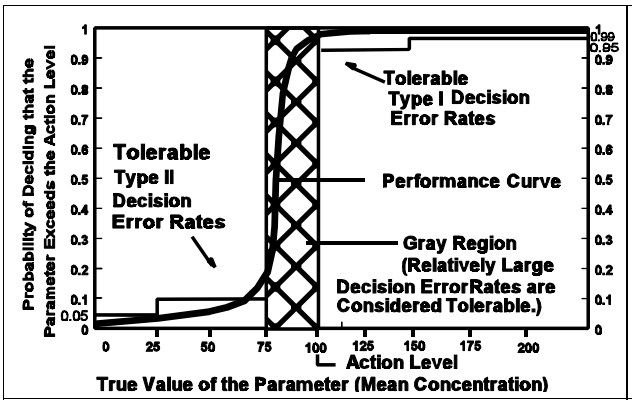

Figure B.2(a) — Decision performance goal diagram null hypothesis: the parameter exceeds the action level.

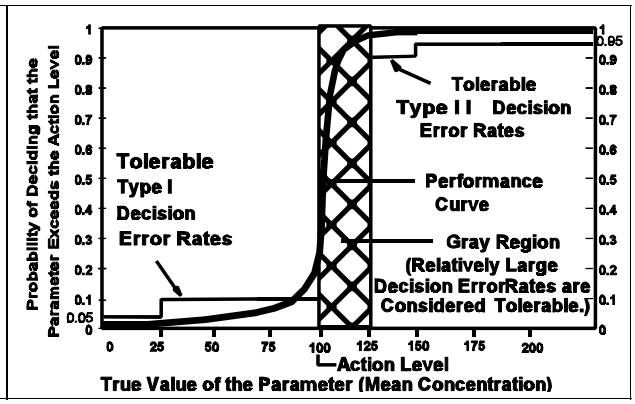

Figure B.2(b) — Decision performance goal diagram null hypothesis: the parameter is less than the action level.

In Figure B.2(b), the default assumption (null hypothesis) was established as "the measured concentration is less than the action level." The Type I error is understood as taking an action when in fact that action is *not* required (e.g., analyte concentrations are really below the action level). The desired limit on the probability of making a Type I error is set at 5 percent if the true concentration is less than 25, and at 10 percent if the true concentration is between 25 and 100. The Type II error is understood as making a decision not to take action to solve an environmental problem (e.g., to remediate) when that action was in fact required (e.g., analyte concentrations are really above an action level). The desired limit on the probability of making a Type II error is set at 10 percent if the true concentrations is between 125 and 150 and at 5 percent if the true concentrations is over 150.

The output of Step 6 is:

- The project planning team's quantitative measure of tolerable decision error rates based on consideration of project resources.

PROCEDURE FOR SPECIFYING LIMITS ON DECISION ERRORS—AN EXAMPLE

Decisionmakers are interested in knowing the true state of some parameter for which action may be proposed. In Step 5 of the DQO process, the parameter, the action level, and the alternative actions were specified in a decision rule. But, decisionmakers cannot positively know the true state because there will always be the potential for uncertainty in estimating the parameter from data. There will be sampling uncertainty, due to spatial and temporal variability in concentrations across the site and from one sample to the next. There will also be analytical measurement uncertainty due to the variability in the measurement process itself. Since it is impossible to eliminate uncertainty, basing decisions on measurement data opens the possibility of making a decision error. Recognizing that decision errors are possible because of uncertainty is the first step in controlling them.

As an example problem, suppose that a decision must be made about whether or not a particular survey unit at a site meets established criteria for residual radioactivity concentrations. Table B.1(a) shows the two possible decision errors that might occur in deciding whether or not a survey unit has been remediated sufficiently so that it may be released. The decision will be based on concentration measurements taken in the survey unit.

As another example problem, suppose that a decision must be made about whether or not a sample contains a particular radionuclide. Table B.1(b) shows the two possible decision errors that might occur in deciding whether or not a sample contains the radionuclide. The decision will be based on a measurement taken on the sample.

TABLE B.1 — Possible decision errors

(a) For survey unit release	
Decision	**True State**
Deciding a survey unit meets the release criterion	when it actually does not
Deciding a survey unit does not meet the release criterion . . .	when it actually does
(b) For radionuclide detection	
Decision	**True State**
Deciding a sample contains the radionuclide	when it actually does not
Deciding a sample does not contain the radionuclide	when it actually does

The probability of making a decision error can be controlled by the use of statistical hypothesis testing. In statistical hypothesis testing, data are used to select between a chosen baseline condition (null hypothesis) and an alternative condition. The test can then be used to decide if there is sufficient evidence to indicate that the baseline condition is unlikely and that the alternative condition is more consistent with the data. Actions appropriate to the alternative conditions would then be appropriate. Otherwise, the default baseline condition remains in place as the basis for decisions and actions. The burden of proof is placed on rejecting the baseline condition. The structure of statistical hypothesis testing maintains the baseline condition as being true until significant evidence is presented to indicate that the baseline condition is not true.

The selection of the baseline condition is important to the outcome of the decision process. The same set of sample data from a survey unit might lead to different decisions depending on what is chosen as the baseline condition.

In deciding if a sample analyzed for a particular radionuclide actually contains that radionuclide, the two possibilities for the baseline condition are:

1) The sample contains the radionuclide, or
2) The sample does not contain the radionuclide.

In this instance, suppose Condition 2, the sample does not contain the radionuclide, is taken as the baseline.[1] The measurement result must be high in order to dismiss the assumption that the sample does not contain the radionuclide. If the measurement is high enough, it is no longer credible that the sample does not contain the radionuclide. Therefore it will be decided that the sample does contain the radionuclide. The framework of statistical hypothesis testing allows one to quantify what is meant by "high enough" and "no longer credible." The measurement value that is considered just "high enough" that the baseline is "no longer credible" is called the "critical value." The baseline condition is called the "null hypothesis," usually denoted $H_{0.}$ The alternate condition is called the alternative hypothesis, usually denoted H_1 or H_A.

Note that if a poor measurement is made—for example, if the sample containing a concentration near the minimum detectable concentration (MDC) is not counted as long as specified in the standard operating procedures—it will be less likely that a result that is clearly above the variability in the measurement of a blank sample will be obtained. Thus, it will be less likely that a sample with a concentration of the radionuclide near the MDC will be detected with greater than the 95 percent probability that is usually specified in MDC calculations. This is another consequence of the structure of statistical hypothesis testing that maintains the baseline condition until convincing evidence is found to the contrary. Poor or insufficient data often will result in the null hypothesis being retained even when it is not true.

In choosing the baseline condition, it is usually prudent to consider which condition will cause the least harm if it is the one that is acted upon, even if it is not true. This is because the baseline will continue to be assumed true unless the data are clearly in conflict with it.

In deciding if a survey unit meets the release criteria for a particular radionuclide, the two possibilities for the baseline condition are:

1) The survey unit does not meet the release criteria, or
2) The survey unit meets the release criteria.

Condition 1 is usually taken as the baseline. This means that the measurement result must be low in order to dismiss the assumption that the survey unit does not meet the release criteria. If the measurement is low enough, it is no longer credible. Therefore it will be decided that the survey unit does meet the release criteria. Again, the framework of statistical hypothesis testing allows one to quantify what is meant by "low enough" and "no longer credible." The null hypothesis, $H_{0.}$, is that the survey unit does not meet the release criteria; the alternative hypothesis, H_A, is the survey unit does meet the release criteria. By phrasing the null hypothesis this way, the benefit of performing a good survey is that it will be more likely that a survey unit that *should* be released

[1] Condition 1 could only be used if it were phrased in reference to a particular concentration, e.g. the sample contains the radionuclide in concentration in excess of *x* pCi/g. Condition 2 implies a concentration of zero.

will be released. On the other hand, a poor survey will generally result in retaining the assumption that the release criterion has not been met even if it has. This arrangement provides the proper incentive for good survey work.

The term "Type I error" is assigned to the decision error made by concluding the null hypothesis is not true, when it actually is true. The term "Type II error" is assigned to the decision error made by concluding the null hypothesis is true, when it actually is not true. The possibility of a decision error can never be totally eliminated, but it can be controlled.

When the decision is to be based on comparing the average of a number of measurements from samples taken over some specified area, sampling uncertainty can be reduced by collecting a larger number of samples. Measurement uncertainty can be reduced by analyzing individual samples several times or using more precise laboratory methods. Which uncertainty is more effective to control depends on their relative magnitude. For much environmental work, controlling the sampling uncertainty error by increasing the number of field samples is usually more effective than controlling measurement uncertainty by repeated radiochemical analyses.

One thing is certain, however, that reducing decision errors requires the expenditure of more resources. Drastically controlling decision error probabilities to extremely small values may be unnecessary for making a reasonable decision. If the consequences of a decision error are minor, a reasonable decision might be made based on relatively crude data. On the other hand, if the consequences of a decision error are severe, sampling and measurement uncertainty should be controlled as much as reasonably possible. How much is enough? It is up to the decisionmaker and the planning team to decide how much control is enough. They must specify tolerable limits on the probabilities for decision errors. If necessary, efforts to reduce sampling and measurement uncertainty to meet these specified limits can then be investigated.

Throughout the remainder of this example, the decision to be made is going to be based on comparing the average of a number of measurements from samples taken over a specific area to a pre-determined limit. The goal of the decisionmaker and planning team is to design a sampling plan that controls the chance of making a decision error to a tolerable level. The strategy outlined below can be used to specify limits on decision errors:

I. Determine the potential range of the parameter of interest.

II. Choose the null hypothesis and identify the Type I and Type II decision errors.

III. Specify a range of concentrations where the consequences of decision errors are relatively minor.

IV. Assign tolerable decision error rates outside of the range specified in III.

I. DETERMINE POTENTIAL RANGE OF THE PARAMETER OF INTEREST

Establish the range of average concentrations likely to be encountered in the survey unit. One must have some idea of the concentration range in order to specify the type of analysis to be done and the sensitivity it must have. It is also the starting point for deciding what differences in concentration are important to detect.

In the example shown in Figure B.3, the project planning team considers a range of feasible concentrations for the radionuclide to be between 0–50 pCi/g. This is based on prior experience of the site, scoping, characterization, and remediation-control survey data.

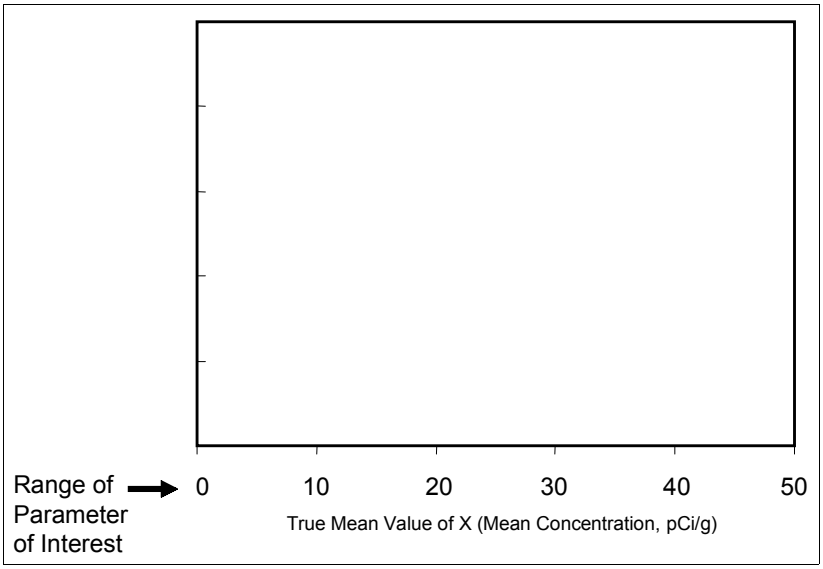

FIGURE B.3 — Plot is made showing the range of the parameter of interest on the x-axis

II. CHOOSE THE NULL HYPOTHESIS AND IDENTIFY DECISION ERRORS

The decision rule states that the action level will be 30 pCi/g for the radionuclide. The project planning team states the null hypothesis as—

H_0: The survey unit concentration exceeds the action level.

The corresponding decision errors are defined as in Table B.2.

TABLE B.2 — Example of possible decision errors with null hypothesis that the average concentration in a survey unit is above the action level

Decision	True State	Consequences	Probability
Deciding a survey unit is below the action level...	...when it actually is above the action level (H_0).	**Type I error**	**α**
Deciding a survey unit is above the action level...	...when it actually is below the action level (H_A).	**Type II error**	**β**

Now that a null hypothesis has been chosen, the meaning of a Type I and a Type II decision error is also defined. In Figure B.4, a line is added showing the action level. A Type I error occurs when the null hypothesis is incorrectly rejected. This means that it is decided that a survey unit with a true mean concentration above the action level may be released. This is the only kind of decision error that can occur if the true concentration is at or above the action level. A Type II error occurs when the null hypothesis is *not* rejected when it is *false*. This means that it is decided that a survey unit with a true mean concentration below the action level may not be released. This is the only kind of decision error that can occur if the true concentration is below the action level. The type of decision error possible at a given value of the true concentration is shown, and a y-axis for displaying control limits on making decision errors, once they have been specified by the project planning team, are also shown in Figure B.4.

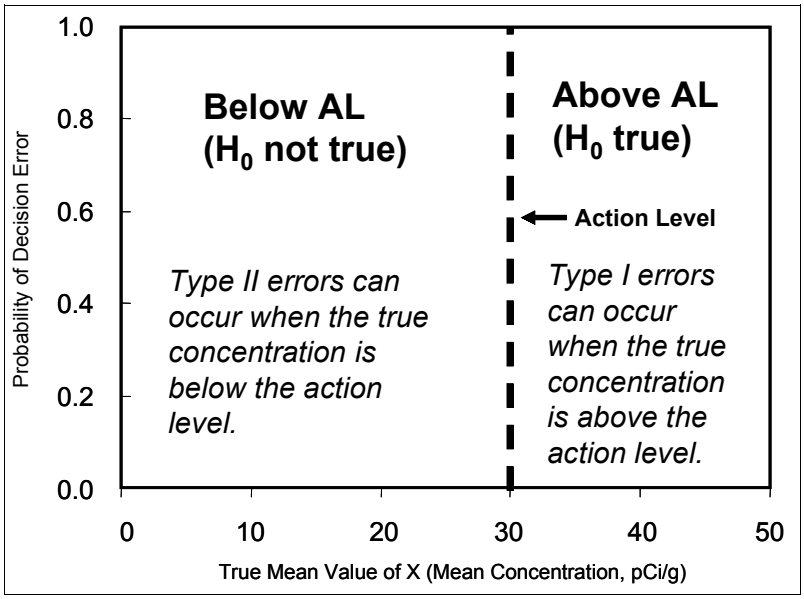

FIGURE B.4 — A line showing the action level, the type of decision error possible at a given value of the true concentration, and a y-axis showing the acceptable limits on making a decision error have been added to Figure B.3

III. SPECIFY A RANGE OF CONCENTRATIONS WHERE THE CONSEQUENCES OF DECISION ERRORS ARE RELATIVELY MINOR

The gray region, or region of uncertainty, indicates an area where the consequences of a Type II decision error are relatively minor. It may not be reasonable to attempt to control decision errors within the gray area. The resources expended to distinguish small differences in concentration could well exceed the costs associated with making the decision error.

In this example, the question is whether it would really make a major difference in the action taken if the concentration is called 30 pCi/g when the true value is 26 or even 22 pCi/g. If not, the gray region might extend from 20 to 30 pCi/g . This is shown in Figure B.5.

The width of the gray region reflects the decisionmaker's concern for Type II decision errors near the action level. The decisionmaker should establish the gray region by balancing the resources needed to "make a close call" versus the consequences of making a Type II decision error. The cost of collecting data sufficient to distinguish small differences in concentration could exceed the cost of making a decision error. This is especially true if the consequences of the error are

judged to be minor.

There is one instance where the consequences of a Type II decision error might be considered major. That is when expensive remediation actions could be required that are not necessary to protect public health. It could be argued that this is always the case when the true concentration is less than the action level. On the other hand, it can be also be argued that remediation of concentrations near, even though not above the action level, will still carry some benefit. To resolve the issue, however, the project planning team knows that not all values of the average concentration below the action level are equally likely to exist in the survey unit. Usually, there is some knowledge, if only approximate, of what the average value of the concentration in the survey unit is. This information can be used to set the width of the gray region. If the planning team is fairly confident that the concentration is less than 20 pCi/g but probably more than 10 pCi/g, they would be concerned about making Type II errors when the true concentration is between 10 and 20 pCi/g. However, they will be much less concerned about making Type II errors when the true concentration is between 20 and 30 pCi/g. This is simply because they do not believe that the true concentration is likely to be in that range. Figure B.6 shows three possible ways that the project planning team might decide to set the gray region. In "A" the project planning team believes the true concentration remaining in the survey unit is about 15 pCi/g, in "B" they believe it to be about 20 pCi/g, and in "C" about 25 pCi/g. In each case, they are less concerned about a decision error involving a true concentration greater than what is estimated to actually remain. They have used

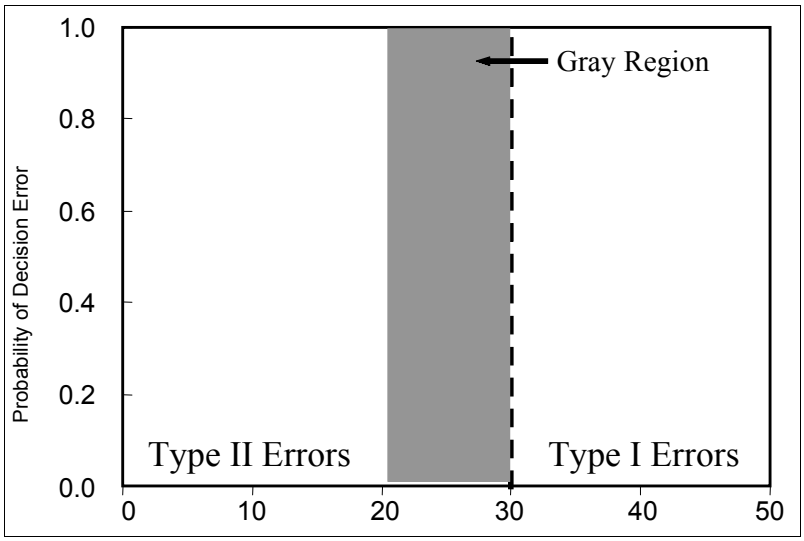

FIGURE B.5 — The gray region is a specified range of values of the true concentration where the consequences of a decision error are considered to be relatively minor

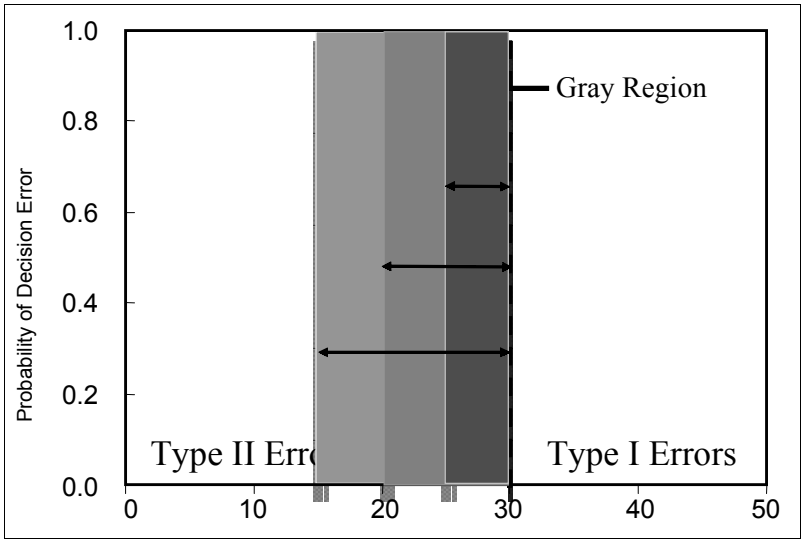

FIGURE B.6 — Three possible ways of setting the gray region. In (A) the project planning team believes the true concentration remaining in the survey unit is about 15 pCi/g, in (B) about 20 pCi/g and in (C) about 25 pCi/g

their knowledge of the survey unit to choose the range of concentration where it is appropriate to expend resources to control the Type II decision error rate. The action level, where further remediation would be necessary, defines the upper bound of the gray region where the probability of a Type I error should be limited. The lower bound of the gray region defines the concentration below which remediation should not be necessary. Therefore, it defines where the probability of a Type II error that would require such an action should be limited.[2]

IV. Assign Tolerable Probability Values for the Occurrence of Decision Errors Outside of the Range Specified in III

As part of the DQO process, the decisionmaker and planning team must work together to identify possible consequences for each type of decision error. Based on this evaluation, desired limits on the probabilities for making decision errors are set over specific concentration ranges. The risk associated with a decision error will generally be more severe as the value of the concentration moves further from the gray region. The tolerance for Type I errors will decrease as the concentration increases. Conversely, the tolerance for Type II errors will decrease as the concentration deceases.

In the example, the decisionmaker has identified 20–30 pCi/g as the area where the consequences of a Type II decision error would be relatively minor. This is the gray region. The tolerable limits on Type I decision errors should be smallest for cases where the decisionmaker has the greatest concern for making an incorrect decision. This will generally be at relatively high values of the true concentration, well above the action level. Suppose, in the example, that the decisionmaker is determined to be nearly 99 percent sure that the correct decision is made, namely, *not* to reject the null hypothesis, *not* to release the survey unit, if the true concentration of radionuclide *X* is 40 pCi/g or more. That means the decisionmaker is only willing to accept a Type I error rate of roughly 1 percent, or making an incorrect decision 1 out of 100 times at this concentration level. This is shown in Figure B.7(a).

If the true concentration of *X* is closer to the action level, but still above it, the decisionmaker wants to make the right decision, but the consequences of an incorrect decision are not considered as severe at concentrations between 30 and 40 pCi/g as they are when the concentration is over 40 pCi/g. The project planning team wants the correct action to be taken at least 90 percent of the time. They will accept an error rate not worse than about 10 percent. They will only accept a data collection plan that limits the potential to incorrectly decide not to take action when it is actually needed to about 1 in 10 times. This is shown in Figure B.7(b).

[2] Had the null hypothesis been chosen differently, the ranges of true concentration where Type I and Type II errors occur would have been reversed.

The decisionmaker and project planning team are also concerned about wasting resources by cleaning up sites that do not represent any substantial risk. Limits of tolerable probability are set low for extreme Type II errors, i.e. failing to release a survey unit when the true concentration is far below the gray region and the action level. They want to limit the chances of deciding to take action when it really is not needed to about 1 in 20 if the true concentration is less than 10 pCi/g. This is shown in Figure B.7(c).

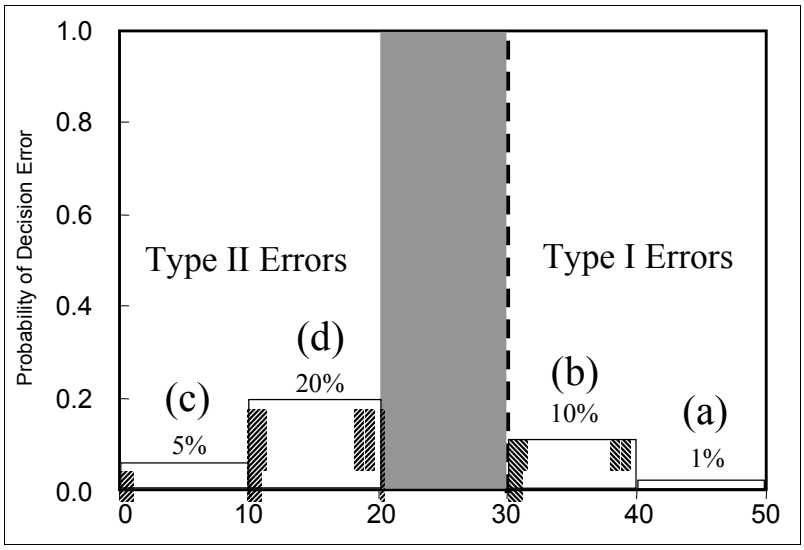

FIGURE B.7 — Example decision performance goal diagram

They are more willing to accept higher decision error rates for concentrations nearer to the gray region. After all, there is some residual risk that will be avoided even though the concentration is below the action level. A Type II error probability limit of 20 percent in the 10–20 pCi/g range is agreed upon. They consider this to be an appropriate transition between a range of concentrations where Type II errors are of great concern (<10 pCi/g) to a range where Type II errors are of little concern. The latter is, by definition, the gray region, which is 20–30 pCi/g in this case . The chance of taking action when it is not needed within the range 10–20 pCi/g is set at roughly 1 in 5. This is shown in Figure B.7(d).

Once the limits on both types of decision error rates have been specified, the information can be displayed on a decision performance goal diagram, as shown in Figure B.7, or made into a decision error limits table, as shown in Table B.3. Both are valuable tools for visualizing and evaluating proposed limits for decision errors.

TABLE B.3 — Example decision error limits table

True Concentration	Correct Decision	Tolerable Probability of Making a Decision Error
0 – 10 pCi/g	Does not exceed	5%
10 – 20 pCi/g	Does not exceed	20%
20 – 30 pCi/g	Does not exceed	gray region: decision error probabilities not controlled
30 – 40 pCi/g	Does exceed	10%
40 – 50 pCi/g	Does exceed	1%

There are no fixed rules for identifying at what level the decisionmaker and project planning team should be willing to tolerate the probability of decision errors. As a guideline, as the possible true values of the parameter of interest move closer to the action level, the tolerance for decision errors usually increases. As the severity of the consequences of a decision error increases, the tolerance decreases.

The ultimate goal of the DQO process is to identify the most resource-effective study design that provides the type, quantity, and quality of data needed to support defensible decisionmaking. The decisionmaker and planning team must evaluate design options and select the one that provides the best balance between cost and the ability to meet the stated DQOs.

A statistical tool known as an estimated power curve can be extremely useful when investigating the performance of alternative survey designs. The probability that the null hypothesis *is* rejected when it *should* be rejected is called the statistical power of a hypothesis test. It is equal to one minus the probability of a Type II error ($1-\beta$). In the example, the null hypothesis is false whenever the true concentration is less than the action level. Figure B.8 shows the power diagram constructed from Figure B.7 by replacing the desired limits on Type II error probabilities, β, with the power, $1-\beta$. The desired limits on Type I error probabilities, α, are carried over without modification, as is the gray region. Drawing a smooth decreasing function through the desired limits results in the desired power curve. A decision performance goal diagram with an estimated power curve can help the project planning team visually identify information about a proposed study design.

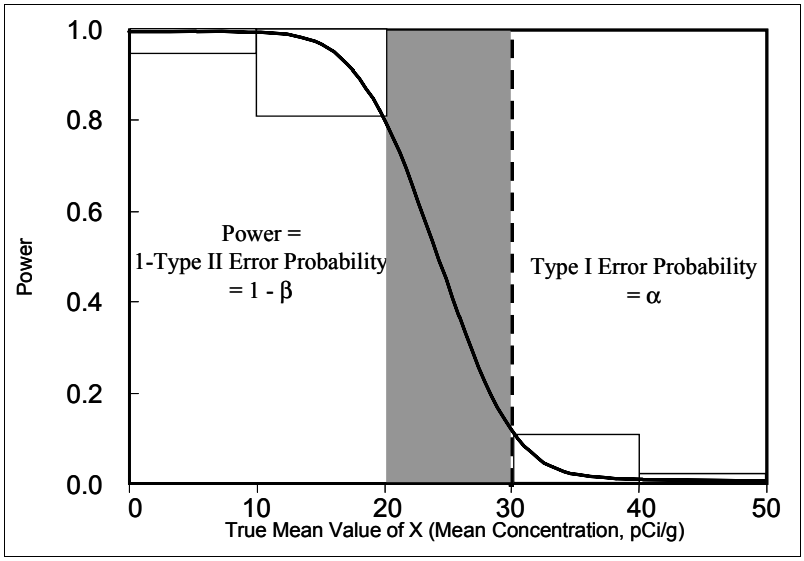

FIGURE B.8 — A power curve constructed from the decision performance goal diagram in Figure B.7

Statisticians can determine the number of measurements needed for a proposed survey design from four values identified on the decision performance goal diagram:

(1) The tolerable limit for the probability of making Type I decision errors, α, at the action level AL).

(2) The tolerable limit for the probability of making Type II decision errors, β, along the

lower bound of the gray region (LBGR).

(3) The width of the gray region, Δ = AL – LBGR, where the consequences of Type II decision errors are relatively minor.

(4) The statistical expression for the total expected variability of the measurement data in the survey unit, σ.

The actual power curve for the statistical hypothesis test can be calculated using these values, and can be compared to the desired limits on the probability of decision errors.

The estimated number of measurements required for a proposed survey design depends heavily on the expected variability of the measurement data in the survey unit, σ. This may not always be easy to estimate from the information available. However, the impact of varying this parameter on the study design is fairly easy to determine during the planning process. Examining a range of reasonable values for σ may not result in great differences in survey design. If so, then a crude estimate for σ is sufficient. If not, the estimate for σ may need to be refined, perhaps by a pilot study of 20 to 30 samples. If the change in the number of samples (due to refining the estimate of σ) is also about 20 to 30 in a single survey unit, it may be better to simply use a conservative estimate of σ that leads to the larger number of samples rather than conduct a pilot study to obtain a more accurate estimate of σ . On the other hand, if several or many similar survey units will be subject to the same design, a pilot study may be worthwhile.

The example in Figure B.9 shows that the probability of making a decision error for any value of the true concentration can be determined at any point on the power curve. At 25 pCi/g, the probability of a Type II error is roughly 45–50 percent. At 35 pCi/g, the probability of a Type I error is roughly 3 percent.

The larger the number of samples required to meet the stated DQOs, the greater the costs of sampling and analysis for a proposed plan. Specifying a narrow gray region and/or very small limits on decision error probabilities indicate a high level of certainty is needed and a larger number of samples will be required.

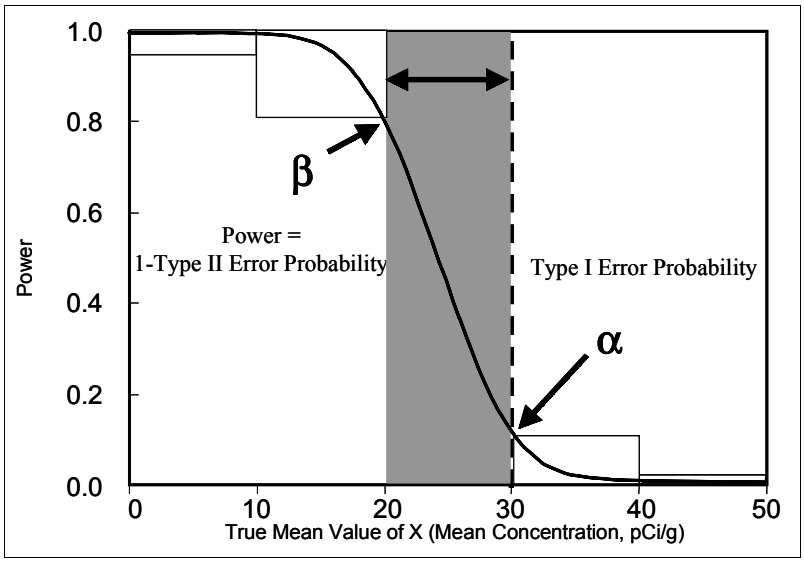

FIGURE B.9 — Example power curve showing the key parameters used to determine the appropriate number of samples to take in the survey unit

Specifying a wide gray region and/or larger limits on decision error probabilities indicates a lower level of certainty is required. A smaller number of samples will be necessary. The required level of certainty should be consistent with the consequences of making decision errors balanced against the cost in numbers of samples to achieve that level of certainty.

If a proposed survey design fails to meet the DQOs within constraints, the decisionmaker and planning team may need to consider:

- ADJUSTING THE ACCEPTABLE DECISION ERROR RATES. For example, the decisionmaker may be unsure what probabilities of decision error are acceptable. Beginning with extremely stringent decision error limits with low risk of making a decision error may require an extremely large number of samples at a prohibitive cost. After reconsidering the potential consequences of each type of decision error, the decisionmaker and planning team may be able to relax the tolerable rates.

- ADJUST THE WIDTH OF THE GRAY REGION. Generally, an efficient design will result when the relative shift, Δ/σ, lies between the values of 1 and 3. A narrow gray region usually means that the proposed survey design will require a large number of samples to meet the specified DQOs. By increasing the number of samples, the chances of making a Type II decision error is reduced, but the potential costs have increased. The wider the gray region, the less stringent the DQOs. Fewer samples will be required, costs will be reduced but the chances of making a Type II decision error have increased. The relative shift, Δ/σ, depends on the width of the gray region, Δ, and also on the estimated data variability, σ. Better estimates of either or both may lead to a more efficient survey design. In some cases it may be advantageous to try to reduce σ by using more precise measurement methods or by forming more spatially homogeneous survey units, i.e. adjusting the physical boundaries of the survey units so that the anticipated concentrations are more homogeneous with them.

B.3.8 DQO Process Step 7: Optimize the Design for Obtaining Data

By the start of Step 7, the project planning team has established their priority of concerns, the definition of the problem, the decision or outcome to address the posed problem, the inputs and boundaries, and the tolerable decision error rates. They have also agreed on decision rules that incorporate all this information into a logic statement about what action to take in response to the decision. During Step 7, the hard decisions are made between the planning team's desire to have measurements with greater certainty and the reality of the associated resource needs (time, cost, etc.) for obtaining that certainty. Another viewpoint of this process is illustrated in Attachment B1. The application of this process to MDC calculations is given in Attachment B2.

During Step 7, the project planning team optimizes the sampling and analytical design and establishes the measurement quality objectives (MQOs) so the resulting data will meet all the established constraints in the most resource-effective manner. The goal is to determine the most

efficient design (combination of sample type, sample number and analytical procedures) to meet all the constraints established in the previous steps. Once the technical specialists and the rest of the project planning team come to agreement about the sampling and analysis design, the operational details and theoretical assumptions of the selected design should be documented.

If a proposed design cannot be developed to meet the limits on decision error rates within budget or other constraints, then the project planning team will have to consider relaxing the error tolerance, adjusting the width of the gray region, redefining the scale of decision, or committing more funding. There is always a trade off among quality, cost, and time. The project planning team will need to develop a consensus on how to balance resources and data quality. If the proposed design requires analysis using analytical protocols not readily available, the project planning team must consider the resources (time and cost) required to develop and validate a method, generate method detection limits relevant to media of concern, and develop appropriate QA/QC procedures and criteria (Chapter 6, *Selection and Application of an Analytical Method*).

If the project entails a preliminary investigation of a site or material for which little is known, the planners may choose to employ MQOs and requirements that typically are achieved by the selected sampling and analytical procedures. At this early point in the project, the lack of detailed knowledge of the site or material may postpone the need for the extra cost of more expensive sampling and analytical procedures and large numbers of samples, until more site or material knowledge is acquired. The less-demanding MQOs, however, should be adequate to further define the site or material. For situations when the measured values are distant from an action level the MQO-compliant data could also be sufficient to support the project decision.

The planning of data collection activities typically is undertaken to determine if a characteristic of an area or item does or does not exist above an action level. Since the area of interest (population) is usually too large to be submitted to analyses, in its entirety, these data collection activities generally include sampling. If sampling is done correctly, the field sample or set of field samples will represent the characteristics of interest and, if analyzed properly, the information gleaned from the samples can be used to make decisions about the larger area. However, if errors occur during implementation of the project, the samples and associated data may not accurately reflect the material from which the samples were collected and incorrect decisions could be made.

The planning team attempts to anticipate, quantify, and minimize the uncertainty in decisions resulting from imprecision, bias, and blunders—in other words, attempts to manage uncertainty by managing its sources. The effort expended in managing uncertainty is project dependent and depends upon what constitutes an acceptable level of decision uncertainty and the proximity of the data to a decision point. For example, Figure B.10(a) presents a situation where the data have significant variability. Yet the variability of the data does not materially add to the uncertainty of the decision since the measurements are so far removed from the action level. More resources could be expended to control the variability. However, the additional expenditure would be unnecessary, since they would not alter the decision or measurably increase confidence in the decision.

In contrast, Figure B.10(b) depicts data with relatively little variability, yet this level of variability is significant since the measured data are adjacent to the action level, which results in increased uncertainty in the decision. Depending upon the consequences of an incorrect decision, it may be advisable to expend more resources with the intention of increasing confidence in the decision.

The outputs of Step 7 are:

- The most resource-effective design for sampling and analysis that will obtain the specific amount and quality of data needed to resolve the problem within the defined constraints; and

- Detailed plans and criteria for data assessment.

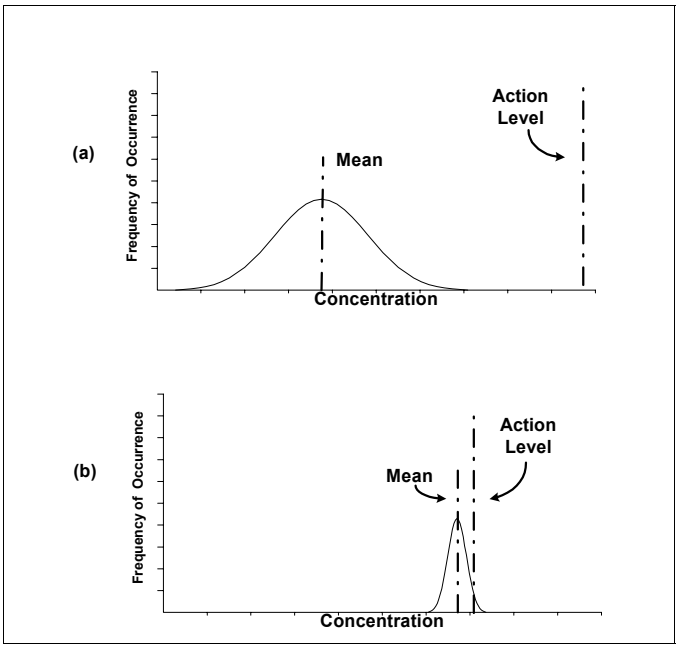

Figure B.10 — How proximity to the action level determines what is an acceptable level of uncertainty

B.4 References

American National Standards Institute (ANSI) N42.23. *American National Standard Measurement and Associated Instrument Quality Assurance for Radioassay Laboratories*. 2003.

American Society for Testing and Materials (ASTM) D5283. *Standard Practice for Generation of Environmental Data Related to Waste Management Activities: Quality Assurance and Quality Control Planning and Implementation*. 1992.

American Society for Testing and Materials (ASTM) D5792. *Standard Practice for Generation of Environmental Data Related to Waste Management Activities: Development of Data Quality Objectives*, 1995.

American Society for Testing and Materials (ASTM) D6051. *Standard Guide for Composite Sampling and Field Subsampling for Environmental Waste Management Activities.* 1996.

Bottrell, D., S. Blacker, and D. Goodman. 1996a. "Application of Decision Theory Methods to the Data Quality Objectives Process." *In*, Proceedings of the Computing in Environmental Resource Management Conference, Air and Waste Management Association.

Bottrell, D., N. Wentworth, S. Blacker, and D. Goodman. 1996b. "Improvements to Specifying Limits on Decision Errors in the Data Quality Objectives Process." *In*, Proceedings of the Computing in Environmental Resource Management Conference, Air and Waste Management Association.

U.S. Environmental Protection Agency (EPA). 1986. *Development of Data Quality Objectives, Description of Stages I and II*. Washington, DC.

U.S. Environmental Protection Agency (EPA). 2000. *Guidance for the Data Quality Objective Process* (EPA QA/G-4). EPA/600/R-96/055, Washington, DC. Available at www.epa.gov/quality1/qa_docs.html.

MARSSIM. 2000. *Multi-Agency Radiation Survey and Site Investigation Manual, Revision 1*. NUREG-1575 Rev 1, EPA 402-R-97-016 Rev1, DOE/EH-0624 Rev1. August. Available at www.epa.gov/radiation/marssim/.

U.S. Nuclear Regulatory Commission (NRC). 1998. *A Nonparametric Statistical Methodology for the Design and Analysis of Final Status Decommissioning Surveys*. NUREG-1505, Rev. 1.

ATTACHMENT B1
Decision Error Rates and the Gray Region for Decisions About Mean Concentrations

B1.1 Introduction

This attachment presents additional information on decision error rates and the gray region. The project planning team will need to specify a range of possible values where the consequences of decision errors are relatively minor—the "gray region." Specifying a gray region is necessary because variability in the population and imprecision in the measurement system combine to produce variability in the data such that the decision may be "too close to call" when the true value is very near the action level. The gray region establishes the minimum separation from the action level, where it is most important that the project planning team control Type II errors.

B1.2 The Region of Interest

The first step in constructing the gray region is setting the range of concentrations that is a region of interest (a range of possible values). This normally means defining the lowest and highest average concentrations at which the contaminant is expected to exist. Usually there is an action level (such as the derived concentration guideline level, DCGL, a regulatory limit) that should not be exceeded. If the project planning team wants a method to measure sample concentrations around this level, they would not select one that worked at concentrations at 10 to 100 times the action level, nor would they select one that worked from zero to half the action level. They would want a method that worked well around the action level—perhaps from 0.1 to 10 times the action level, or from one-half to two times the action level. For the purpose of the example in this attachment, the action level is 1.0 and the project planning team selected a region of interest that is zero to twice the action level (0–2), as shown on the x-axis in Figure B1.1.

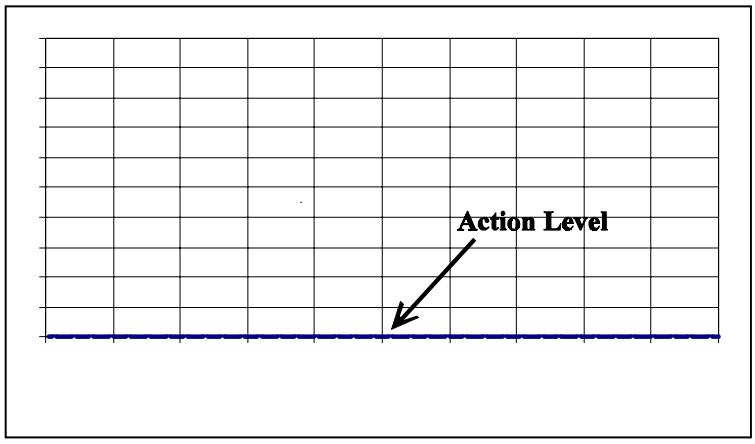

FIGURE B1.1 — The action level is 1.0

B1.3 Measurement Uncertainty at the Action Level

The action level marks the concentration level that the project planning team must be able to distinguish. The project planning team wants to be able to tell if the measured concentration is

above or below the action level. Does this mean that the project planning team needs to be able to distinguish 0.9999 times the action level from 1.0001 times the action level? Sometimes, but not usually. This is fortunate, because current measurement techniques are probably not good enough to distinguish that small a difference in concentrations.

How close to the action level can the project planning team plan to measure? This example assumes that the standard uncertainty (1 sigma, σ) of the measured concentration is 10 percent of the action level. With that kind of measurement "precision," can the project planning team tell the difference between a sample with 0.9 times the action level from one right at the action level? Not always. Figure B1.2 shows the distribution of the concentration that is measured (assuming a normal distribution). This means that about 16 percent of the time, the measured concentration (in the shaded area) will appear to be 0.9 times the action level or less, even though the true concentration is exactly equal to the action level.

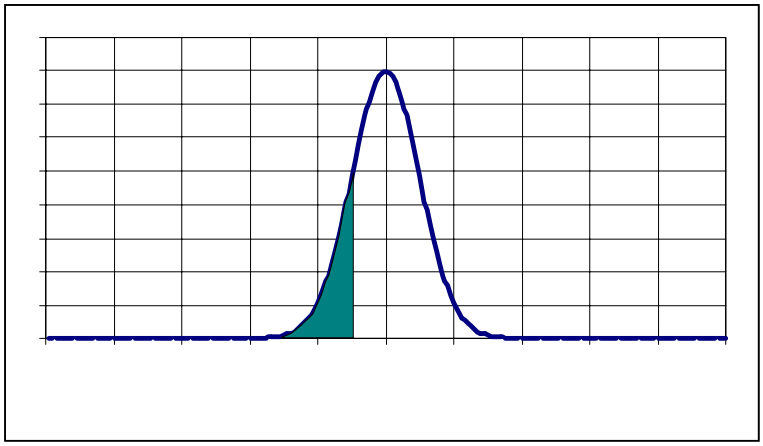

FIGURE B1.2 — The true mean concentration is 1.0. The standard uncertainty of the distribution of measured concentrations is 0.1.

Similarly, about 16 percent of the time, the measured concentration will appear to be at or above the action level (as shown in the shaded area in Figure B1.3), even though the true concentration is only 0.9 times the action level.

The problem is, when there is only the measurement result to go by, the project planning team cannot tell the difference with confidence. If the measured concentration is 0.9, it is more likely that the true concentration is 0.9 than it is 1.0, but there remains a chance that it is really 1.0. The moral of the story is that measurement variability causes some ambiguity about what the true concentration is. This translates into some uncertainty in the decisionmaking process. This uncertainty can be controlled with careful planning, but it can never be eliminated. On the other hand, the ambiguity caused by measurement variability really only affects the ability to distinguish between concentrations that are "close together." In our example, 0.9 and 1.0 are "close together" not because 0.1 is a small difference, but because there is a great degree of overlap between the curves shown in Figures B1.2 and B1.3. The peaks of the two curves are separated by 0.1, but each curve spreads out over a value several times this amount on both sides. The most common statistical measure of the amount of this spread is the standard deviation. The standard deviation in this case is 0.1, the same as the amount of separation between the peaks. If the peaks were separated by 0.3, i.e. 3 standard deviations, there would be far less overlap, and

far less ambiguity. There would be very little uncertainty in deciding which curve a single measurement belonged to, and consequently whether the mean was 0.7 or 1.0.

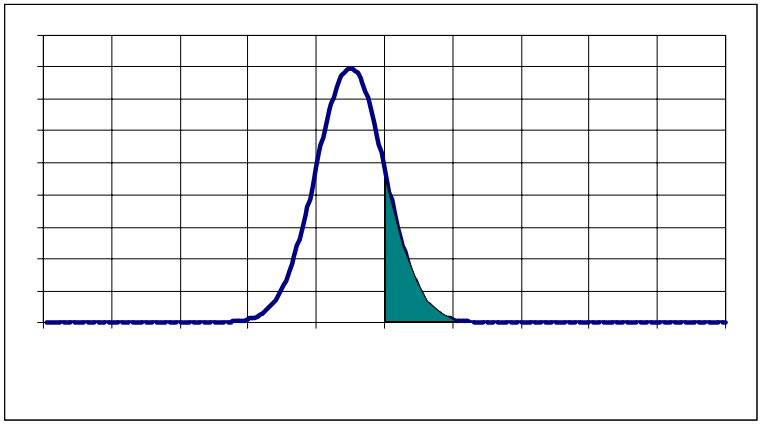

FIGURE B1.3 — The true mean concentration is 0.9. The standard uncertainty of the distribution of measured concentrations is 0.1.

From this discussion, at least two very important conclusions can be drawn:

(1) True mean concentrations that are "very close together" are not easily distinguished by a single measurement.

(2) A useful way for determining what is meant by "very close together" is by measuring the separation in concentration in standard deviation units. Concentrations that are one or fewer standard deviations apart are close together, whereas concentrations that are three or more standard deviations apart are well separated.

From conclusion (1), it is immediately apparent that no matter how small the measurement variability is, there must be some separation between the concentration values to be distinguished. It is not possible to determine whether or not the concentration is on one side or the other of "a bright line" (e.g. above or below the action level). Instead, one must be content to pick two concentrations separated by a finite amount and attempt to tell them apart. These two concentrations define what is known as the gray region, because one cannot be certain about deciding whether concentrations that lie between the two boundaries are above or below the action level. To illustrate this with the example, if the measured concentration is 0.95—exactly in the middle of the gray region between the two concentrations to be distinguished—it is equally likely that the true concentration is 0.9 as it is 1.0 (Figure B1.4).

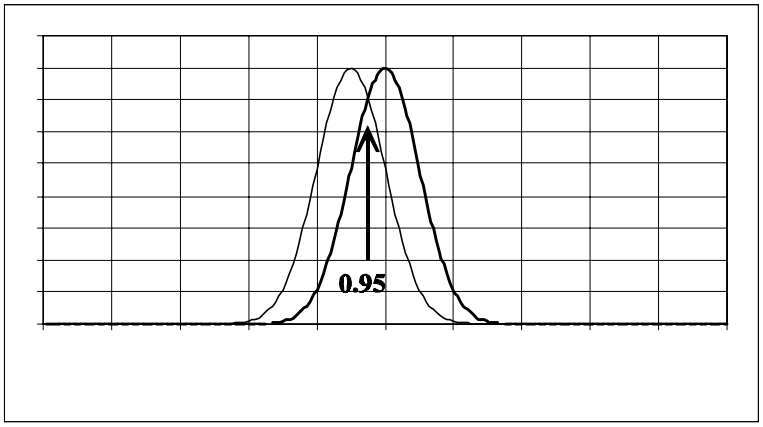

FIGURE B1.4 — If 0.95 is measured, is the true mean concentration 1.0 (right) or 0.9 (left)? The standard uncertainty of the distribution of measured concentrations is 0.1.

To formalize this process of distinguishing whether the true concentration is above our upper bound or below our lower bound,

two hypotheses will be defined and a statistical hypothesis test will be used to decide between the two.

B1.4 The Null Hypothesis

How does the project planning team decide whether the true concentration is above or below the gray region? By formulating hypotheses. Suppose it has been decided that it is important to distinguish whether the true mean concentration is above 1.0 or below 0.9. These concentrations then correspond to the "upper bound of the gray region" (UBGR) and to the "lower bound of the gray region" (LBGR), respectively.

The project planning team starts by asking which mistake is worse: (1) deciding the true concentration is less than the action level when it is actually above, or (2) deciding the true concentration is above the action level when it is actually below?

Mistake (1) may result in an increased risk to human health in the general population following site release, while mistake (2) may result in increased occupational risks or a waste of resources that might have been used to reduce risks elsewhere.

The way to avoid the "worse mistake" is to assume the worse case is true, i.e., make the worse case the baseline or null hypothesis. For example, to avoid mistake (1), deciding the true concentration is less than the action level when it is actually above, the null hypothesis should be that the true concentration is above the action level. Only when the data provide convincing evidence to the contrary will it be decided that the true concentration is less than the action level. Borderline cases will default to retaining (not rejecting) the null hypothesis.

Note that while the null hypothesis must be, in fact, either true or false, the data cannot prove that it is true or false with absolute certainty. When the probability of obtaining the given data is sufficiently low under the conditions specified by the null hypothesis, it is evidence to decide that the null hypothesis should be rejected. On the other hand, if the null hypothesis is not rejected, it is not the same as proving that the null hypothesis is true. It only means that there was not enough evidence, based on the probability of observing the data obtained, to decide to reject it.

Notice that in Figure B.2 (Section B.3.7 on page B-11), the risk that is elevated in the gray region is that of making a Type II error. That is, in the gray region, the Type II error rate exceeds the tolerable limit set at the boundary of the gray region. The Type I error rate remains fixed. (It is fixed at exactly the value used to determine the critical value for the statistical test.) A Type II error is incorrectly accepting (failing to reject) the null hypothesis when it is false. So another way to think about choosing the null hypothesis is to decide which mistake is less tolerable, and framing the null hypothesis so that kind of mistake corresponds to a Type I error (i.e., incorrectly rejecting the null hypothesis when it is actually true).

Another pragmatic consideration is that the project planning team really does not want to make a mistake that is likely to remain undiscovered or will be difficult or expensive to correct if it is discovered. If the project planning team decides the true concentration is less than the action level, the team is not likely to look at the data again. That would mean that the mistake would probably not be discovered until much later (e.g. during a confirmatory survey), if at all. On the other hand, if the project planning team decides that the true concentration is over the action level when it really is not, the project planning team will discover the mistake while they are trying to figure out how to take action (i.e., to remediate). This is a pragmatic reason to set the null hypothesis so as to assume the true concentration exceeds the action level. This null hypothesis will not be rejected unless the project planning team is certain that the true concentration is below the action level. This way of choosing the null hypothesis will not work when the action level is so low compared with the expected data variability that no reasonable values of Type II error rates can be achieved. This can occur, for example, when the action level is close to (or even equal to) zero. In that case, if the action level is chosen to be the UBGR, the lower bound might have to be negative. It is impossible to demonstrate that the true concentration is less than some negative value, because negative concentrations are not possible. In such cases, there may be no alternative but to choose as the null hypothesis that the action level is met. Then a concentration that is unacceptably higher than the action level is chosen for the UBGR.

CASE 1: ASSUME THE TRUE CONCENTRATION IS OVER 1.0

If a true concentration of 1.0 or more is over a regulatory limit, such as a DCGL, the project planning team will not want to make mistake (1) above. So they generally will choose as the null hypothesis that the true concentration exceeds the action level of 1.0. How sure does the project planning team need to be? To be 95 percent sure, they would have to stay with their assumption that the true concentration is over 1.0 unless the measured concentration is 0.84 or less (Figure B1.5). The project planning team knows that they will only observe a concentration less than 0.84 about 5 percent of the time when the true concentration is really 1.0. That is, the measurement has to be less than 0.84 to be 95 percent sure the true concentration is less than 1.0. This is an example of a *decision rule* being used to decide between two alternative hypotheses. If a concentration of less than 0.84 is observed, one can decide that the true concentration is less than 1.0—

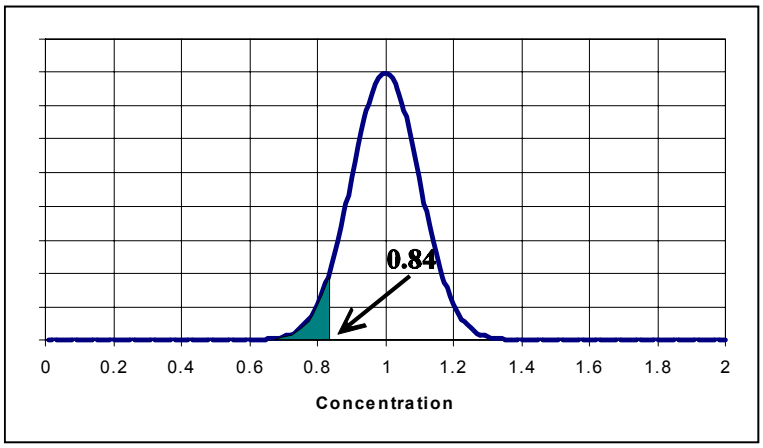

FIGURE B1.5 — When the true mean concentration is 1.0, and the standard uncertainty of the distribution of measured concentrations is 0.1, a measured concentration of 0.84 or less will be observed only about 5 percent of the time

i.e., the null hypothesis is rejected. Otherwise, if a concentration over 0.84 is observed, there is not enough evidence to reject the null hypothesis, and one retains the assumption that the true concentration is over 1.0.

But what if the true concentration is 0.9 or less? Under the null hypothesis, how often will the project planning team say that the true concentration is over 1.0 when it is really only 0.84? As seen in Figure B1.6, there is only a 50-50 chance of making the right decision when the true concentration really is 0.84. That is the price of being sure the action level is not exceeded. The Type II error rate, when the true concentration is 0.9, is over 50 percent.

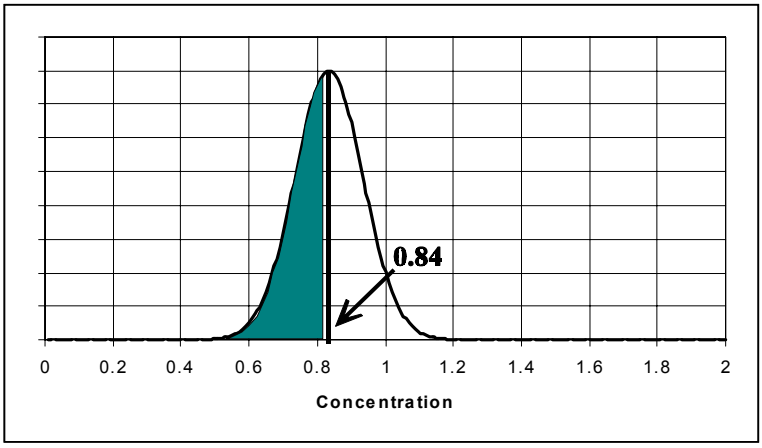

FIGURE B1.6 — When the true mean concentration is 0.84, and the standard uncertainty of the distribution of measured concentrations is 0.1, a measured concentration of 0.84 or less will be observed only about half the time

How low does the true concentration have to be in order to have a pretty good chance of deciding that the true concentration is below the limit? To be 95 percent sure, the true concentration needs to be twice as far below the action level as the decision point (i.e., critical value), namely at about 0.68. That is, the project planning team will need a concentration of 0.68 or less to be 95 percent sure that they will be able to decide the true concentration is less than 1.0 (see the unshaded portion in Figure B1.7). The "critical value" (or decision point) is the measured value that divides the measurement results into two different sets: (1) those values that will cause the null hypothesis to be rejected and (2) those values that will leave the null hypothesis as the default. In other words, it is only when the true concentration is 0.68 or less that the project planning team can be pretty sure that they will decide the true concentration is less than 1.0. Notice that the project planning team could change the decision rule. For example, they could decide that if the measured concentration is less than 0.9, they will reject the null hypothesis. Examining Figures B1.2

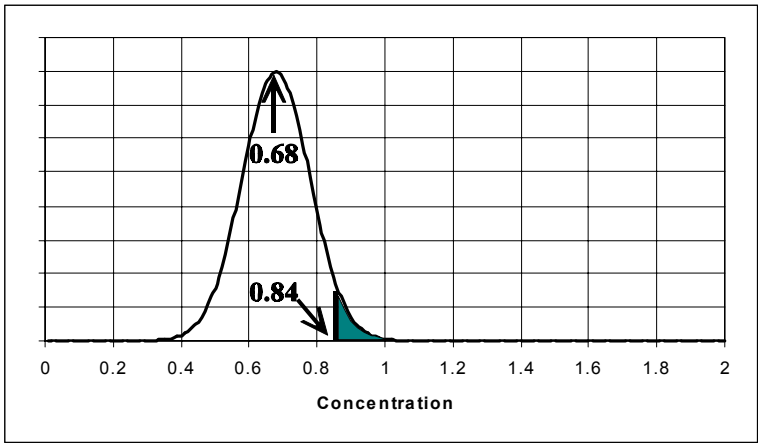

FIGURE B1.7 — When the true mean concentration is 0.68 and the standard uncertainty of the distribution of measured concentrations is 0.1, a measured concentration over 0.84 will be observed only about 5 percent of the time

and B1.3 once again, the Type I error rate will be about 16 percent instead of 5 percent. However, the Type II error rate will decrease from 50 percent to 16 percent. Fortunately, by moving the decision point—called the "critical value"—the error rates can be adjusted. However, reducing one error rate necessarily increases the other. The only way to decrease both decision error rates is to reduce the uncertainty (standard deviation) of the distribution of measured concentrations.

CASE 2: ASSUME THE TRUE CONCENTRATION IS 0.9

As stated previously, the mistake that is most serious determines the null hypothesis. Suppose that the project planning team determined that it is worse to decide that the true concentration is over 1.0 when it is 0.9 (than it is to decide it is 0.9 when it is 1.0). Then, the default assumption (the null hypothesis) would be that the true concentration is less than 0.9, unless the measured concentration is large enough to convince the planning team otherwise. Using a decision rule (critical value) of 1.06, the planning team can decide the true concentration is over 1.0 with only a 5 percent chance that it is actually 0.9 or less (Figure B1.8). The team will have to have a true concentration of 1.22 or more to be 95 percent sure that they will be able to decide the true concentration is over 1.0.

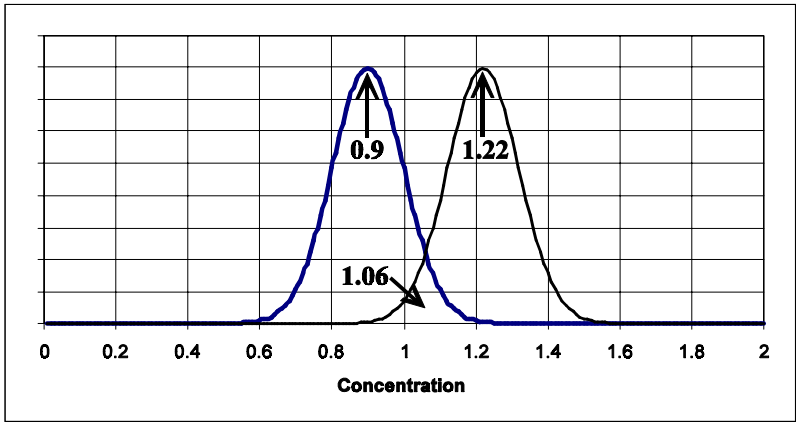

Figure B1.8 — The true mean concentration is 0.9 (left) and 1.22 (right). The standard uncertainty of the distribution of measured concentrations is 0.1.

B1.5 The Gray Region

In the previous sections of this attachment, the project planning team:

- Set the region of interest for the measured concentrations between zero and about twice the action level;
- Assumed that the true concentration exceeds 1.0, unless they measure "significantly" below that, the default assumption (null hypothesis);
- Defined "significantly below" to mean a concentration that would be observed less than 5 percent of the time, when the true concentration is actually 1.0. To describe their uncertainty, the project planning team used the normal distribution, with a relative standard deviation of 10 percent at the action level, as a model;
- Developed an operational decision rule: If the measured concentration is less than 0.84, then

decide the true concentration is less than 1.0. Otherwise, decide there is not enough reason to change the default assumption (null hypothesis); and

- Found using this operational decision rule that they were pretty sure (95 percent) of deciding that the true concentration is less than 1.0 only when the true concentration is actually 0.68 or less.

If the true concentration is between 0.68 and 1.0, all the project planning team really can say is that the probability of correctly deciding that the true concentration is less than 1.0 will be between 5 percent (when the true concentration is just under 1.0) and 95 percent (when the true concentration is 0.68). In other words, when the true concentration is in the range of 0.68 to 1.0, the probability of incorrectly deciding that the true concentration is not less than 1.0 (i.e., the probability of making a Type II error) will be between 5 percent (when the true concentration is 0.68) and 95 percent (when the true concentration is just under 1.0). This range of concentrations, 0.68 to 1.0, is the "gray region."

When the null hypothesis is that the true concentration exceeds the action level (1.0), the gray region is bounded from above by the action level. This is where α (the desired limit on the Type I error rate) is set. It is bounded from below at the concentration where β (the desired limit on the Type II error rate) is set. There is some flexibility in setting the LBGR. If the project planning team specifies a concentration, they can calculate the probability β. If they specify β, they can calculate the value of the true concentration that will be correctly detected as being below 1.0 with probability $1-\beta$.

Often it will make sense to set the LBGR at a concentration at, or slightly above, the project planning team's best estimate of the true concentration based on all of the information that is available to them. Then the width of the gray region will truly represent the minimum separation in concentration that it is important to detect, namely, that between the action level and what actually is believed to be there.

In our example, the project planning team found that they needed the true concentration to be 0.68 or less to be at least 95 percent sure that they will correctly decide (by observing a measured value of 0.84 or less) that the true concentration is less than 1.0. If the project planning team is not satisfied with that, the team can find that a true concentration of 0.71 will be correctly detected 90 percent of the time (also by observing a measured value of 0.84 or less). The critical value, or decision point, is determined by α, not β.

If the project planning team decides to raise the LBGR (i.e., narrow the gray region) the Type II error rate at the LBGR goes up. If they lower the LBGR (i.e., widen the gray region) the Type II error rate at the LBGR goes down. Nothing substantive is really happening. The project planning team is merely specifying the ability to detect that the null hypothesis is false (i.e., reject the null hypothesis because it is not true) at a particular concentration below the action level called the

LBGR.

If the project planning team wants to make a substantive change, they need to change the probability that an error is made. That is, they need to change the uncertainty (standard deviation) of the measurements. Suppose the relative standard deviation of the measurements at the action level is 5 percent instead of 10 percent. Then the value of the true concentration that will be correctly detected to be below the action level (by observing a measured value of 0.92 or less) 95 percent of the time, is 0.84. Cutting the standard deviation of the measurement in half has cut the (absolute) width of the gray region in half, but left the width of the gray region in standard deviations unchanged. Previously, with $\sigma = 10$ percent, the width of the gray region was $1.0 - 0.68 = 0.32 = 3.2\ (0.10) = 3.2\sigma$. As Figure B1.9 illustrates, with $\sigma = 5$ percent, the width of the gray region is $1.0 - 0.84 = 0.16 = 3.2\ (0.05) = 3.2\sigma$.

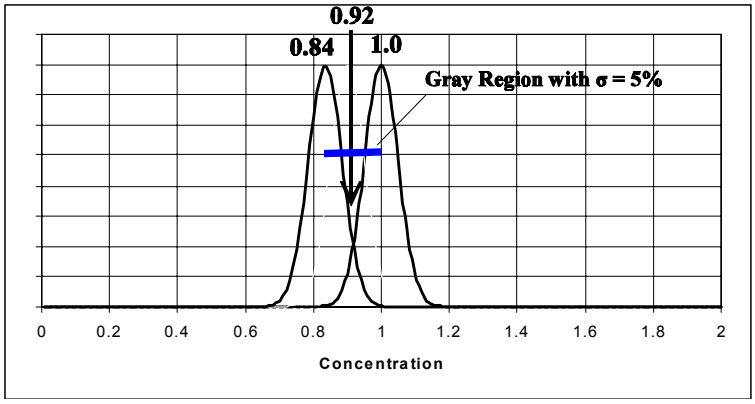

FIGURE B1.9 — The true mean concentration is 0.84 (left) and 1.0 (right). The standard uncertainty of the distribution of measured concentrations is 0.05. The relative shift is 3.2.

What is important is the width of the gray region in standard deviations; not the width of the gray region in concentration. In order to achieve the same specified Type II error rate at the LBGR, the action level and the LBGR must be separated by the same number of standard deviations. The width of the gray region (action level minus LBGR) will be denoted by delta (Δ), the "shift." Δ/σ is how many standard deviations wide the gray region is. Δ/σ is called the "relative shift."

If the gray region is less than one standard deviation wide, the Type II error rate may be high at the LBGR. The only way to improve the situation would be to decrease the standard deviation (i.e., increase the relative shift, Δ/σ). This can be done by employing a more precise measurement method or by averaging several measurements. When the width of the gray region is larger than about three standard deviations (i.e., Δ/σ exceeds 3), it may be possible to use a simpler, less expensive measurement method or take fewer samples. Unnecessary effort should not be expended to achieve values of Δ/σ greater than 3.

B1.6 Summary

The mistake that is "worse" defines the null hypothesis and also defines a "Type I" error. The probability of a Type I error happening is called the "Type I error rate," and is denoted by alpha (α). Under the original null hypothesis (Case 1: Assume the true concentration is over 1.0), a

Type I error would be deciding that the concentration was less than 1.0 when it really was not. In general, a Type I error is deciding against the null hypothesis when it is actually true. (A Type I error is also called a "false positive." This can be confusing when the null hypothesis appears to be a "positive" statement. Therefore, MARLAP uses the neutral terminology.)

The "less serious" mistake is called a Type II error, and the probability of it happening is the "Type II error rate," denoted by beta (β). Under the original null hypothesis that the concentration was 1.0 or more, a Type II error would be deciding that the concentration was more than 1.0 when it really was not. In general, a Type II error is not deciding against the null hypothesis when it is actually false.

In both Case 1 and Case 2, the probability of both Type I errors and Type II errors were set to 5 percent. The probabilities were calculated at multiples of the standard deviation, assuming a normal distribution. The data may not always be well described by a normal distribution, so a different probability distribution may be used. However, the probability of a Type I error is always calculated as the probability that the decisionmaker will reject the null hypothesis when it is actually true. This is simple enough, as long as there is a clear boundary for the parameter of interest.

The parameter of interest in both Case 1 and Case 2 was the true concentration. The true concentration had a limit of 1.0. Therefore, all the project planning team had to do was calculate the probability that they would get a measured concentration that would cause them to decide that the true concentration was less than 1.0, even though it was equal to 1.0. In the example, the project planning team actually started with the probability (5 percent) and worked out the critical value. The "critical value" (or decision point) is the measured value that divides the measurement results into two different sets: (1) those values that will cause the null hypothesis to be rejected and (2) those values that will leave the null hypothesis as the default.

The Type I and Type II error rates, α and β, often are both set at 5 percent. This is only by tradition. Neither error rate needs to be set at 5 percent, nor do they have to be equal. The way the project planning team should set the value is by examining the consequences of making a Type I or a Type II error. What consequences will happen as a result of making each type of error? This is a little different than the criterion that was used to define the null hypothesis. It may be that in some circumstances, a Type II error is riskier than a Type I error. In that case, consider making α bigger than β.

ATTACHMENT B2
Decision Error Rates and the Gray Region for Detection Decisions

B2.1 Introduction

This section is provided to present some additional discussion on the subject of applying the Data Quality Objectives (DQO) process to the problem of measurement detection capability. In particular, "not detected" does not mean zero radioactivity concentration. To understand this, one needs to examine the concept of "minimum detectable concentration" (MDC). This involves the DQO process and limiting decision error rates.

B2.2 The DQO Process Applied to the Detection Limit Problem

STEP 1. PROBLEM STATEMENT
To determine if the material that is being measured contains radioactivity.

STEP 2. IDENTIFY THE DECISION
Decide if the material contains radioactivity at a level that requires action.

STEP 3. IDENTIFY INPUTS TO THE DECISION
What level of radioactivity in the material is important to detect?

STEP 4. DEFINE THE STUDY BOUNDARIES
How much material is to be measured, what instrumentation/analysis is available, how much time and resources are available for the measurements.

STEP 5. DEVELOP A DECISION RULE
This is an "if...then" rule that specifies the parameter of interest to be measured, and an action level against which it is compared in order to choose between alternative actions. At this stage, it is assumed that the true value of the parameter can be measured exactly without uncertainty. Such a decision rule in this case might be "If the true concentration in the sample is greater than zero, appropriate action will be taken. Otherwise, no action is required."

STEP 6. SPECIFY LIMITS ON DECISION ERROR RATES
Develop an operational rule so that when the measurement is made, a decision on the appropriate action to take can be made. This rule takes into account that there is uncertainty in any measurement, and therefore there is the possibility of making decision errors. When the material is processed and inserted into an instrument, the measurement is made and the instrument output is a result that is a number. The decision rule involves taking that numerical result and comparing it to a pre-determined number called the critical value. If the

result is greater than the critical value, the decision is made to treat the material as containing radioactivity above the action level, and then taking the appropriate action. The critical value will vary depending on the limits on decision errors rates that are specified.

The material either contains radioactivity or it does not. Unfortunately, it is impossible to determine absolutely whether the material does or does not contain radioactivity. Decisions can only be based on the result of measurements. There are four possibilities:

- The material *does not* contain radioactivity, and the measurement results in a value below the critical value and so it is *decided* that it *does not* contain radioactivity.

- The material *does* contain radioactivity, and the measurement results in a value above the critical value and so it is *decided* that it *does* contain radioactivity.

- The material *does not* contain radioactivity, and the measurement results in a value above the critical value and so it is *decided* that it *does* contain radioactivity. This would be a decision error.

- The material *does* contain radioactivity, and the measurement results in a value below the critical value and so it is *decided* that it *does not* contain radioactivity. This also would be a decision error.

Note that one never knows if a decision error is made, one only knows the result of the measurement. Measurements are not perfect, people make mistakes, and decision errors are unavoidable. However, recognizing that decision errors exist does allow their severity to be controlled. Several steps are necessary in order to create the framework for controlling decision error rates. These are described in the following sections.

B2.3 Establish the Concentration Range of Interest

Step three of the DQO process determined a level of radioactivity concentration in the material that is important to detect. This is also sometimes called an *action level* (such as the DCGL, a regulatory limit) that should not be exceeded. It is also important to define a region of interest ranging from the lowest to the highest average concentrations at which the contaminant is expected to exist. If the project planning team wants a method to measure sample concentrations around the action level, they would not select one that only worked at concentrations at 10 to 100 times the action level, nor would they select one that only worked from zero to half the action level. They would want a method that worked well around the action level—perhaps from 0.1 to 10 times the action level, or from one-half to two times the action level. For the purpose of the example in this attachment, the action level is 1.0 and the project planning team selected a region of interest that is zero to twice the action level (0–2), as shown on the x-axis of Figure B2.1. The

first thing to notice is that Figure B2.1 ranges from −1 to 2 and not 0 to 2. Why is this?

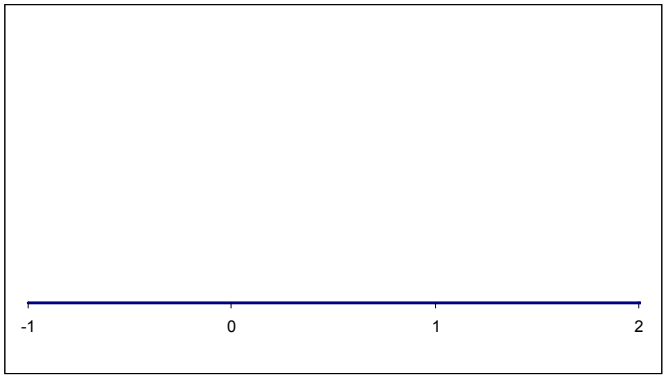

Figure B2.1 — Region of interest for the concentration around the action level of 1.0

If a blank sample is placed in the instrument, the "true concentration" is zero. The instrument will produce a reading that is a number, and not necessarily the same reading each time. This is shown in Figure B2.2(a). Usually, the instrument output must be converted to a concentration value using a calibration factor. For simplicity, this example will assume that the calibration factor is 100, and in the remaining figures the measurement results will be shown directly in concentration. The zero point of concentration is at the average instrument reading when "nothing" (a blank) is being measured. In Figure B2.2(a) this is 100. The distribution of many measurements of nothing will look like Figure B2.2(b). This is obtained from Figure B2.2(a) by subtracting the average blank reading (100) and dividing by the calibration factor (also 100). The spread in these measurement results is characterized by the standard deviation of this distribution. In Figure B2.2(b), the standard deviation is 0.2. For the problem to be actually addressed, the standard deviation may be larger or smaller than this, but it will not be zero. There is always some variability in measurements, and this will always cause some uncertainty about whether or not the decisions based on these measurements are correct.

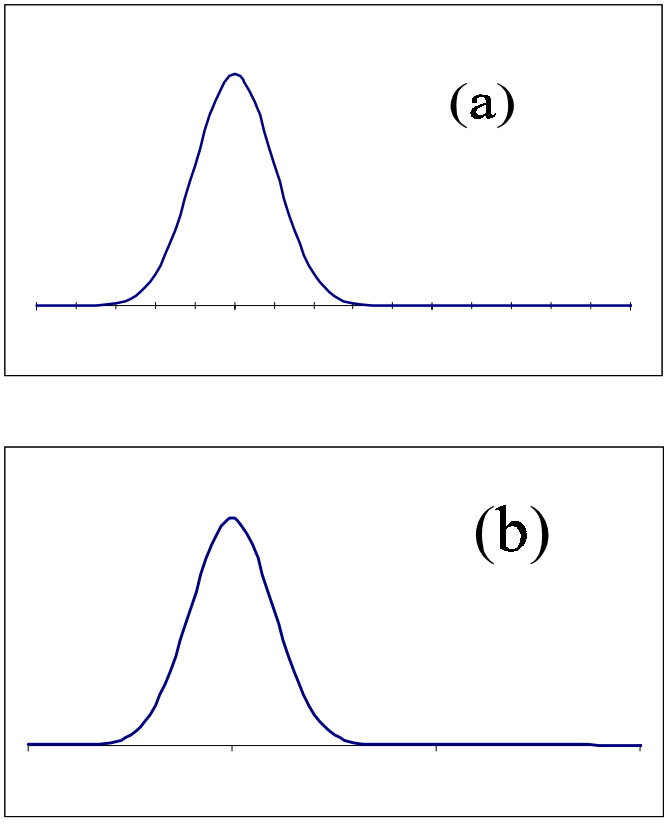

FIGURE B2.2 (a) — The distribution of blank (background) readings. (b) The true concentration is 0.0. The standard deviation of the distribution of measured concentrations is 0.2.

Consider a possible decision rule: Decide that there is radioactivity in the sample if the measurement result is greater than zero. (This means that the "critical value" is zero.)

Figure B2.2 shows that if the critical value for the decision is made equal to zero, a decision that there is radioactivity in the

sample will be made about half the time, even when there is nothing to measure. Also, notice that unless the instrument reading is negative, it is not possible to decide that there is no radioactivity in the sample. There is nothing contradictory about this. The zero point on the x-axis was chosen simply to be the average measurement of "nothing." About half the time a measurement of nothing will be larger, and about half the time it will be smaller. This does not imply anything about concentrations being negative. It is about the variability of measurement readings, not the true concentration.

Decisionmakers might not be too happy about a decision rule that will lead them to the wrong conclusion half of the time. How can this be improved? Notice that if the critical value is made larger, the wrong conclusion (that there is radio-activity when there is none) will be made less often. If the critical value for the example is 0.2, it will be decided that there is radioactivity in the sample when the measurement result is greater than 0.2. From the example in Figure B2.3, this will be estimated to happen about 16 percent of the time.

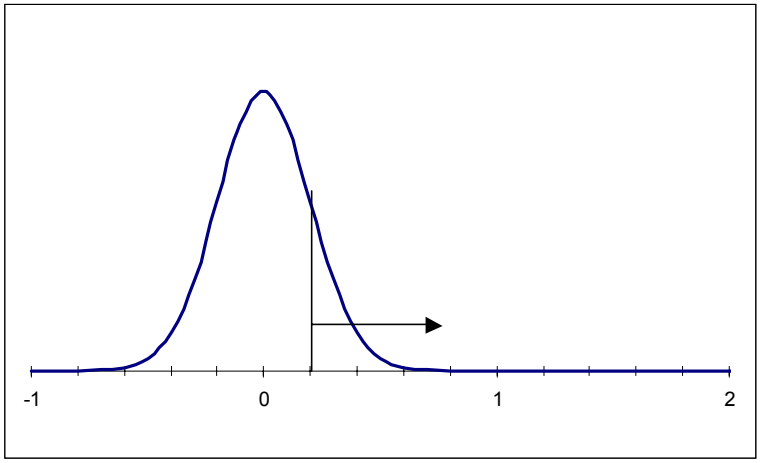

FIGURE B2.3 — The true concentration is 0.0, and the standard deviation of the distribution of measured concentrations is 0.2. A critical value of 0.2 would be exceeded about 16 percent of the time.

By making the critical value larger and larger, the probability can be reduced practically to zero of deciding that there is radioactivity when there is not. This apparently happy solution comes at a price. To see that, just consider the opposite situation. Suppose, instead of "nothing," there is a concentration of 0.2 in the sample (in this example, units are irrelevant). If a sample with this concentration is measured often, the distribution of results might look like Figure B2.4.

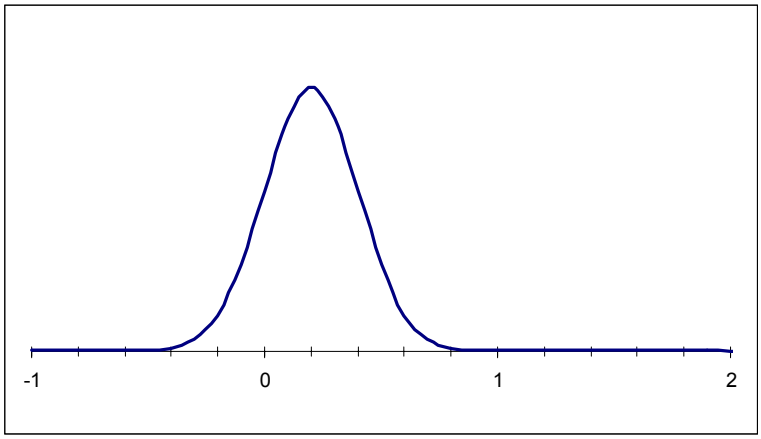

Figure B2.4 — The true concentration is 0.2 and the standard deviation of the distribution of measured concentrations is 0.2

Notice that with a critical value of 0.2, a decision that there is radioactivity in this sample will only be made about half the time. Even if the critical value were zero,

a decision that there is radioactivity in the sample would only be made about 84 percent of the time.

As shown above, there are two types of decision errors that can be made: that there is radioactivity when there is not, or that there is no radioactivity when there is. What the figures show is that by making the critical value for the decision rule bigger, one can reduce the chances of making first kind of decision error, but doing so will increase the chance of making the second kind of decision error. Making the critical value for the decision rule smaller will reduce chances of the second kind of decision error, but will increase the chance of the first kind of decision error.

This example used a measurement variability (standard deviation) of 0.2. What if the variability is larger or smaller? By looking at the figures, one can conclude that *no matter what the variability actually is*:

(1) If a critical value of zero is used, one will conclude that there is radioactivity in a sample that actually contains nothing about half the time.

(2) If a critical value is selected equal to the standard deviation, one will conclude that there is radioactivity in a sample that actually contains nothing about 16 percent of the time. (A slight modification of the figures would show that if the critical value equals two times the standard deviation, one will conclude that there is radioactivity in a sample that actually contains nothing about 2.5 percent of the time.)

(3) If a critical value of zero is used, one will conclude that there is *no* radioactivity in a sample that actually contains a concentration that is numerically equal to the standard deviation about 16 percent of the time. (A slight modification of the figures would show that if the critical value were equal to zero, one will conclude that there is *no* radioactivity in a sample that actually contains a concentration equal to twice the standard deviation about 2.5 percent of the time.)

(4) If a critical value is selected equal to the standard deviation, one will conclude that there is *no* radioactivity in a sample that actually contains a concentration numerically equal to the standard deviation about half the time.

The key is to notice that it is not the numerical value of the variability alone nor the numerical value of the concentration alone, that determines the probability of a decision error. It is the ratio of the concentration to the standard deviation that is important. In essence, the standard deviation determines the *scale* of the x-axis (concentration axis) for this problem. Background determines the zero point of the concentration axis.

The MDC for a measurement process is the concentration that the sample must contain so that

the probability of a Type II decision error is limited to 5 percent. (Other values may be chosen, but 5 percent is most commonly used in this context.)

This means that if a sample containing a concentration equal to the MDC is measured, about 95 percent of the time the measurement result will lead to the decision that the sample contains radioactivity. This is shown in Figure B2.5.

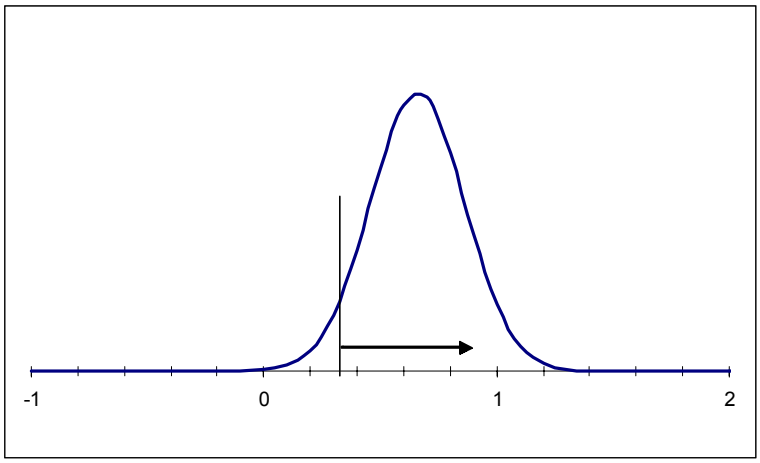

FIGURE B2.5 — The true value of the concentration is 0.66 and the standard deviation of the distribution of measured concentrations is 0.2. A critical value of 0.33 will be exceeded about 95 percent of the time.

However, if a sample containing a blank is measured, the probability that the measurement result will lead to the decision that the sample contains radioactivity will be only about 5 percent. This is shown in Figure B2.6.

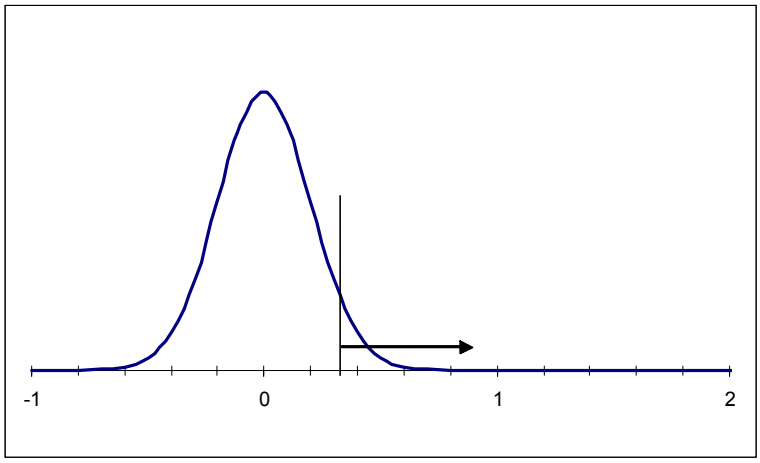

FIGURE B2.6 — The true value of the measured concentration is 0.0 and the standard deviation of the measured concentrations is 0.2. A critical value of 0.33 would be exceeded about 5 percent of the time.

For this example, Figure B2.7 summarizes the relationship among the distribution of measurements on a blank, the critical value, the MDC, and the action level.

The critical value used to limit the decision error of concluding that there is radioactivity in a sample that actually contains a blank to 5 percent, is about 1.5 or 2 times the measurement variability when measuring a blank. Limiting the decision error of concluding that there is no radioactivity in a sample that actually contains a concentration equal to the MDC, results in an MDC that is usually about twice the critical value. Consequently, the MDC is usually about 3 or 4 times the measurement variability when measuring a blank.

B2.4 Estimate the Measurement Variability when Measuring a Blank

The measurement variability when measuring a blank is thus a key parameter for planning. The best way to get a handle on this is by making many measurements of a blank sample and

computing the standard deviation of the measurement results.

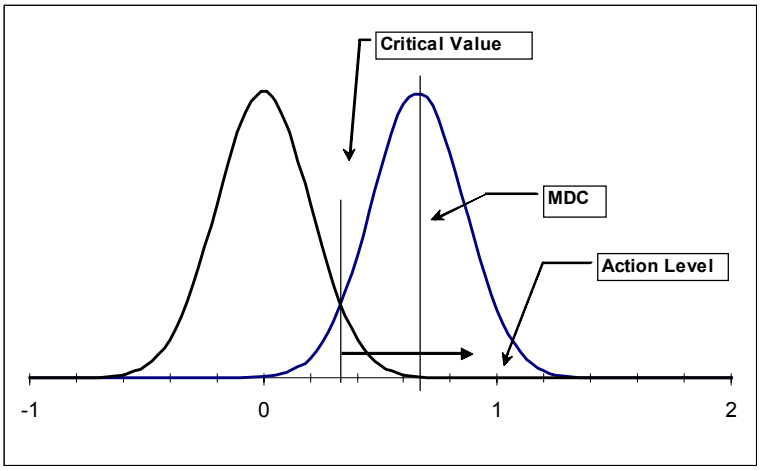

Figure B2.7 — The standard deviation of the normally distributed measured concentrations is 0.2. The critical value is 0.33, the MDC is 0.66 and the action level is 1.0.

What can be concluded about the ability to measure "nothing" (i.e., no analyte)? Radioactivity present at a concentration less than the MDC may be detected, but less than 95 percent of the time. If the "true concentration" is at half the MDC (right at the critical value), the presence of radioactivity will be detected about half the time, and about half the time it will not. Concentrations lower than the critical value will be detected less often. The only way to do better is to reduce the measurement variability. This usually can only be done by either taking more measurements or by using an instrument or measurement process that has less variability when measuring a blank sample.

So what does it mean if a sample is measured, and a decision was made that there was no radioactivity? (This is another way of saying that no radioactivity was detected.) By itself, such a statement means nothing, and has no value unless one knows *the level of radioactivity that could be detected if it were there*—i.e. the MDC.

Similarly, a criterion for action specifying that no radioactivity be detected in a sample must be qualified by information on how hard one must look. That is, the MDC must be specified, which in turn implies a certain limit on the variability of the measurement procedure.

In either case, one can never measure zero. One can only decide from a measurement, with a prescribed limit on the probability of being wrong, that if enough radioactivity were there, it would be found. If it is not found, it does not mean it is not present; it only means that whatever might be there is unlikely to be more than the MDC.

In conclusion, an action level must be determined, and the MDC must be below it. Only then can radioactivity concentrations of concern can be detected with any degree of certainty. Conversely, specifying a measurement process implies an action level (level of concern) that is at or above the MDC.