

## APPENDIX D

### THE PLANNING PHASE OF THE DATA LIFE CYCLE

The planning phase of the Data Life Cycle is carried out using the Data Quality Objectives (DQO) Process. The DQO Process is a series of planning steps based on the scientific method for establishing criteria for data quality and developing survey designs (EPA 1994a, 1987b, 1987c). The level of effort associated with planning is based on the complexity of the survey. Large, complicated sites generally receive a significant amount of effort during the planning phase, while smaller sites may not require as much planning effort.

Planning radiological surveys using the DQO Process can improve the survey effectiveness and efficiency, and thereby the defensibility of decisions. It also can minimize expenditures related to data collection by eliminating unnecessary, duplicative, or overly precise data. The use of the DQO Process assures that the type, quantity, and quality of environmental data used in decision making will be appropriate for the intended application. It provides systematic procedures for defining the criteria that the survey design should satisfy, including when and where to perform measurements, the level of decision errors for the survey, and how many measurements to perform.

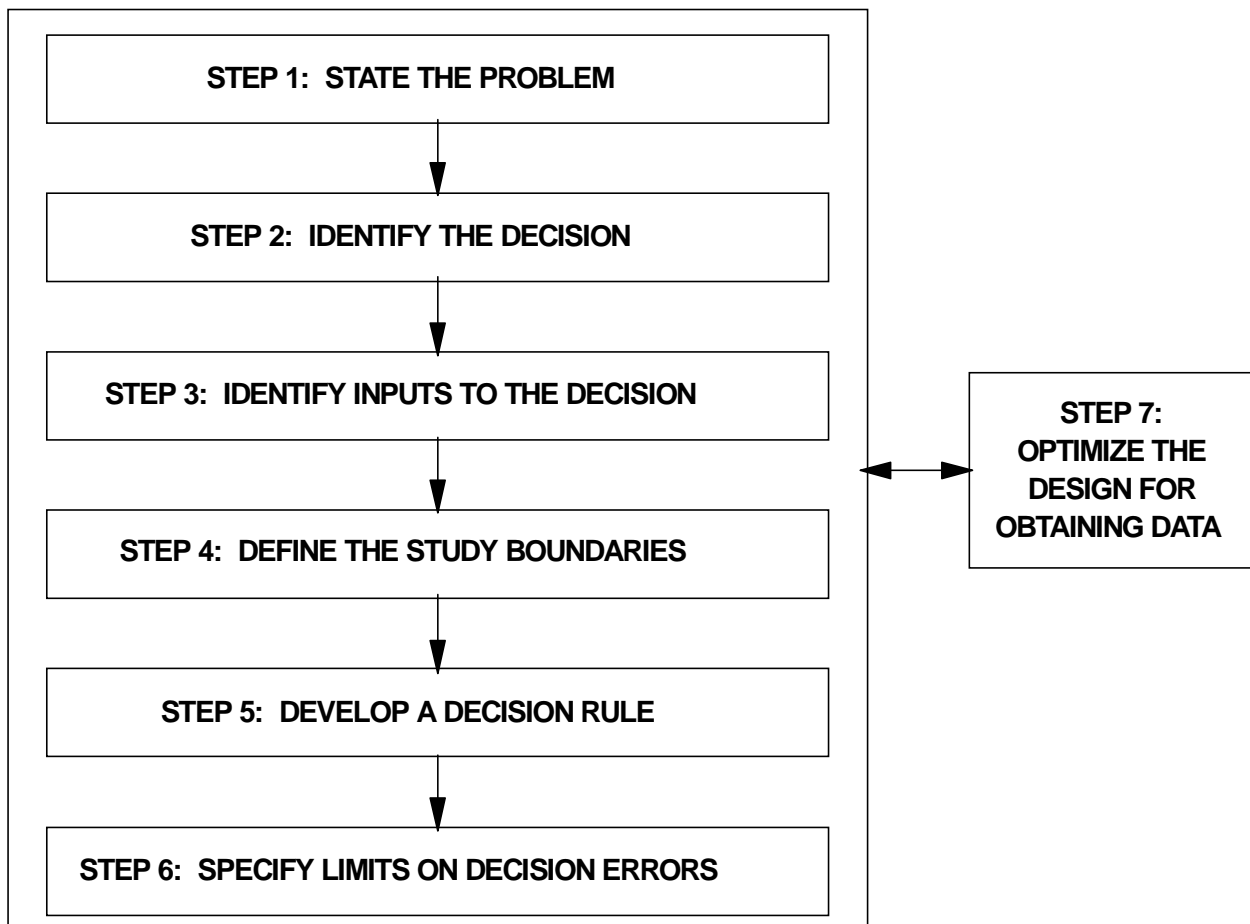
The expected output of planning a survey using the DQO Process is a quality assurance project plan (QAPP). The QAPP integrates all technical and quality aspects of the Data Life Cycle, and defines in detail how specific quality assurance and quality control activities will be implemented during the survey.

The DQO Process provides for early involvement of the decision maker and uses a graded approach to data quality requirements. This graded approach defines data quality requirements according to the type of survey being designed, the risk of making a decision error based on the data collected, and the consequences of making such an error. This approach provides a more effective survey design combined with a basis for judging the usability of the data collected.

DQOs are qualitative and quantitative statements derived from the outputs of the DQO Process that:

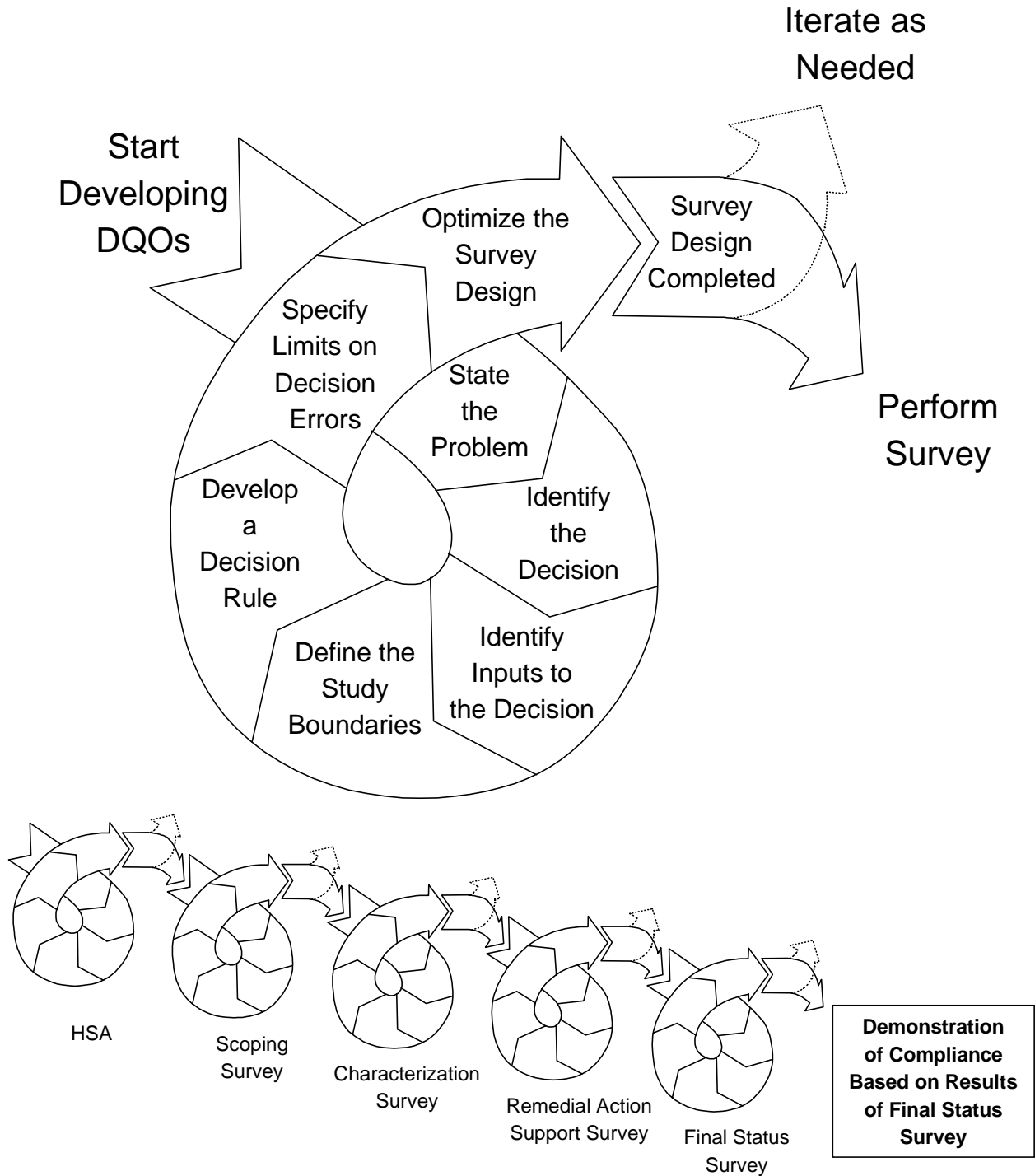
- clarify the study objective
- define the most appropriate type of data to collect
- determine the most appropriate conditions for collecting the data
- specify limits on decision errors which will be used as the basis for establishing the quantity and quality of data needed to support the decision

The DQO Process consists of seven steps, as shown in Figure D.1. The output from each step influences the choices that will be made later in the Process. Even though the DQO Process is depicted as a linear sequence of steps, in practice it is iterative; the outputs of one step may lead to reconsideration of prior steps as illustrated in Figure D.2. For example, defining the survey unit boundaries may lead to classification of the survey unit, with each area or survey unit having a different decision statement. This iteration is encouraged since it ultimately leads to a more efficient survey design. The first six steps of the DQO Process produce the decision performance criteria that are used to develop the survey design. The final step of the Process develops a survey design based on the DQOs. The first six steps should be completed before the final survey design is developed, and every step should be completed before data collection begins.



**Figure D.1 The Data Quality Objectives Process**

When the DQO Process is used to design a survey, it helps ensure that planning is performed properly the first time and establishes measures of performance for the data collector (implementation) and the decision maker (assessment) during subsequent phases of the Data Life Cycle. DQOs provide up-front planning and define decision maker/data collector relationships by presenting a clear statement of the decision maker's needs. This information is recorded in the QAPP.



**Figure D.2 Repeated Applications of the DQO Process Throughout the Radiation Survey and Site Investigation Process**

DQOs for data collection activities describe the overall level of uncertainty that a decision maker is willing to accept for survey results. This uncertainty is used to specify the quality of the measurement data required in terms of objectives for precision, accuracy, representativeness, comparability, and completeness. These objectives are presented in detail in Section 9.3.2 and Appendix N.

The DQO Process is a flexible planning tool that can be used more or less intensively as the situation requires. For surveys that have multiple decisions, such as characterization or final status surveys, the DQO Process can be used repeatedly throughout the performance of the survey. Decisions made early in decommissioning are often preliminary in nature. For this reason, a scoping survey may only require a limited planning and evaluation effort. As the site investigation process nears conclusion the necessity of avoiding a decision error becomes more critical.

The following sections briefly discuss the steps of the DQO Process, especially as they relate to final status survey planning, and list the outputs for each step in the process. The outputs from the DQO Process should be included in the documentation for the survey plan.

## **D.1 State the Problem**

The first step in any decision making process is to define the problem so that the focus of the survey will be unambiguous. Since many sites or facilities present a complex interaction of technical, economic, social, and political factors, the success of a project is critically linked to a complete but uncomplicated definition of the problem .

There are four activities associated with this step:

- identifying members of the planning team and stakeholders
- identifying the primary decision maker or decision-making method
- developing a concise description of the problem
- specifying available resources and relevant deadlines for the study

The expected outputs of this step are:

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

For a final status survey, examples of planning team members and stakeholders are described in Section 3.2. A description of the problem would typically involve the release of all or some portion of a site to demonstrate compliance with a regulation. The resources and deadlines are typically identified on a site-specific basis.

## D.2 Identify the Decision

The goal of this step is to define the question that the survey will attempt to resolve and identify alternative actions that may be taken based on the outcome of the survey. The combination of these two elements is called the decision statement. The decision statement would be different for each type of survey in the Radiation Survey and Site Investigation Process, and would be developed based on the survey objectives described in Chapter 5.

There are four activities associated with this step in the DQO Process:

- identifying the principal study question
- defining the alternative actions that could result from resolution of the principal study question
- combining the principal study question and the alternative actions into a decision statement
- organizing multiple decisions

The expected output from this step is a decision statement that links the principal study question to possible solutions to the problem.

For a final status survey, the principal study question could be: “Is the level of residual radioactivity in the survey units in this portion of the site below the release criterion?” Alternative actions may include further remediation, re-evaluation of the modeling assumptions used to develop the DCGLs, re-assessment of the survey unit to see if it can be released with passive controls, or a decision not to release the survey unit. The decision statement may be: “Determine whether or not all the survey units in this portion of the site satisfy the release criterion.”

## D.3 Identify the Inputs to the Decision

Collecting data or information is necessary to resolve most decision statements. In this step, the planning team focuses on the information needed for the decision and identifies the different types of information needed to resolve the decision statement.

The key activities for this step include:

- Identifying the information required to resolve the decision statement. Ask general questions such as: “Is information on the physical properties of the site required?” or: “Is information on the chemical characteristics of the radionuclide or the matrix required?” Determine which environmental variables or other information are needed to resolve the decision statement.

## Appendix D

- Determining the sources for each item of information. Identify and list the sources for the required information.
- Identifying the information needed to establish the action level or the derived concentration guideline level (DCGL) based on the release criterion. The actual numerical value will be determined in Step 5 (*i.e.*, Section D.5).
- Confirming that appropriate measurement methods exist to provide the necessary data. A list of potentially appropriate measurement techniques should be prepared based on the information requirements determined previously in this step. Field and laboratory measurement techniques for radionuclides are discussed in Chapters 6 and 7 of this manual. Information on using field and laboratory equipment, their detection limits and analytical costs are listed in Appendix H. This performance information will be used in Steps 5 and 7 of the DQO Process.

The expected outputs of this step are:

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

For the final status survey, the list of information inputs generally involves measurements of the radioactive contaminants of concern in each survey unit. These inputs include identifying survey units, classifying survey units, identifying appropriate measurement techniques including measurement costs and detection limits, and whether or not background measurements from a reference area or areas need to be performed. The list of environmental variables measured during the final status survey is typically limited to the level of residual radioactivity in the affected media for each survey unit.

### **D.4 Define the Boundaries of the Study**

During this step the planning team should develop a conceptual model of the site based on existing information collected in Step 1 of the DQO Process or during previous surveys. Conceptual models describe a site or facility and its environs, and present hypotheses regarding the radionuclides present and potential migration pathways. These models may include components from computer models, analytical models, graphic models, and other techniques. Additional data collected during decommissioning are used to expand the conceptual model.

The purpose of this step is to define the spatial and temporal boundaries that will be covered by the decision statement so data can be easily interpreted. These attributes include:

- spatial boundaries that define the physical area under consideration for release (site boundaries)

- spatial boundaries that define the physical area to be studied and locations where measurements could be performed (actual or potential survey unit boundaries)
- temporal boundaries that describe the time frame the study data represents and when measurements should be performed
- spatial and temporal boundaries developed from modeling used to determine DCGLs

There are seven activities associated with this step:

- specifying characteristics that define the true but unknown value of the parameter of interest
- defining the geographic area within which all decisions must apply
- when appropriate, dividing the site into areas or survey units that have relatively homogeneous characteristics
- determining the time frame to which the decision applies
- determining when to collect data
- defining the scale of decision making
- identifying any practical constraints on data collection

The expected outputs of this step are:

- a detailed description of the spatial and temporal boundaries of the problem (a conceptual model)
- any practical constraints that may interfere with the full implementation of the survey design

Specifying the characteristics that define the true but unknown value of the parameter of interest for the final status survey typically involves identifying the radionuclides of concern. If possible, the physical and chemical form of the radionuclides should be described. For example, describing the residual radioactivity in terms of total uranium is not as specific or informative as describing a mixture of uraninite ( $\text{UO}_2$ ) and uranium metaphosphate ( $\text{U}(\text{PO}_3)_4$ ) for natural abundances of  $^{234}\text{U}$ ,  $^{235}\text{U}$ , and  $^{238}\text{U}$ .

As an example, the study boundary may be defined as the property boundary of a facility or, if there is only surface contamination expected at the site, the soil within the property boundary to a depth of 15 cm. When appropriate (typically during and always before final status survey design), the site is subdivided into survey units with relatively homogeneous characteristics based on information collected during previous surveys. The radiological characteristics are defined by the area classification (Class 1, Class 2, or Class 3) while the physical characteristics may include structures vs. land areas, transport routes vs. grassy areas, or soil types with different radionuclide transfer characteristics.

The time frame to which the final status survey decision applies is typically defined by the regulation. For example: “The data are used to reflect the condition of radionuclides leaching into ground water over a period of 1,000 years.” Temporal boundaries may also include seasonal conditions such as winter snow cover or summer drought that affect the accessibility of certain media for measurement.

For the final status survey, the smallest, most appropriate subsets of the site for which decisions will be made are defined as survey units. The size of the survey unit and the measurement frequency within a survey unit are based on classification, site-specific conditions, and relevant decisions used during modeling to determine the DCGLs.

## **D.5 Develop a Decision Rule**

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

There are three activities associated with this step:

- specifying the statistical parameter that characterizes the parameter of interest
- specifying the action level for the study
- combining the outputs of the previous DQO steps into an "if...then..." decision rule that defines the conditions that would cause the decision maker to choose among alternative actions

Certain aspects of the site investigation process, such as the HSA, are not so quantitative that a statistical parameter can be specified. Nevertheless, a decision rule should still be developed that defines the conditions that would cause the decision maker to choose among alternatives.

The expected outputs of this step are:

- the parameter of interest that characterizes the level of residual radioactivity
- the action level
- an “if...then...” statement that defines the conditions that would cause the decision maker to choose among alternative actions

The parameter of interest is a descriptive measure (such as a mean or median) that specifies the characteristic or attribute that the decision maker would like to know about the residual contamination in the survey unit.



The mean is the value that corresponds to the “center” of the distribution in the sense of the “center of gravity” (EPA 1989a). Positive attributes of the mean include: 1) it is useful when the action level is based on long-term, average health effects, 2) it is useful when the population is uniform with relatively small spread, and 3) it generally requires fewer samples than other parameters of interest. Negative attributes include: 1) it is not a very representative measure of central tendency for highly skewed distributions, and 2) it is not useful when a large proportion of the measurements are reported as less than the detection limit (EPA 1994a).

The median is also a value that corresponds to the “center” of a distribution, but where the mean represents the center of gravity the median represents the “middle” value of a distribution. The median is that value such that there are the same number of measurements greater than the median as less than the median. The positive attributes of the median include: 1) it is useful when the action level is based on long-term, average health effects, 2) it provides a more representative measure of central tendency than the mean for skewed populations, 3) it is useful when a large proportion of the measurements are reported as less than the detection limit, and 4) it relies on few statistical assumptions. Negative attributes include: 1) it will not protect against the effects of extreme values, and 2) it is not a very representative measure of central tendency for highly skewed distributions (EPA 1994a).

The nonparametric statistical tests discussed in Chapter 8 are designed to determine whether or not the level of residual activity uniformly distributed throughout the survey unit exceeds the  $DCGL_w$ . Since these methods are based on ranks, the results are generally expressed in terms of the median. When the underlying measurement distribution is symmetric, the mean is equal to the median. The assumption of symmetry is less restrictive than that of normality because the normal distribution is itself symmetric. If, however, the measurement distribution is skewed to the right, the average will generally be greater than the median. In severe cases, the average may exceed the  $DCGL_w$  while the median does not. For this reason, MARSSIM recommends comparing the arithmetic mean of the survey unit data to the  $DCGL_w$  as a first step in the interpretation of the data (see Section 8.2.2.1).

The action level is a measurement threshold value of the parameter of interest that provides the criterion for choosing among alternative actions. MARSSIM uses the investigation level, a radionuclide-specific level of radioactivity based on the release criterion that results in additional investigation when it is exceeded, as an action level. Investigation levels are developed for both the Elevated Measurement Comparison (EMC) using scanning techniques and the statistical tests using direct measurements and samples. Section 5.5.2.6 provides information on investigation levels used in MARSSIM.

The mean concentration of residual radioactivity is the parameter of interest used for making decisions based on the final status survey. The definition of residual radioactivity depends on whether or not the contaminant appears as part of background radioactivity in the reference area. If the radionuclide is not present in background, residual radioactivity is defined as the mean

## Appendix D

concentration in the survey unit. If the radionuclide is present in background, residual radioactivity is defined as the difference between the mean concentration in the survey unit and the mean concentration in the reference area selected to represent background. The term *1-sample case* is used when the radionuclide does not appear in background, because measurements are only made in the survey unit. The term *2-sample case* is used when the radionuclide appears in background, because measurements are made in both the survey unit and the reference area.

Figure D.3 contains a simple, hypothetical example of the 1-sample case. The upper portion of the figure shows a probability distribution of residual radionuclide concentrations in the surface soil of the survey unit. The parameter of interest is the location of the mean of this distribution, represented by the vertical dotted line and denoted by the symbol  $D$ .

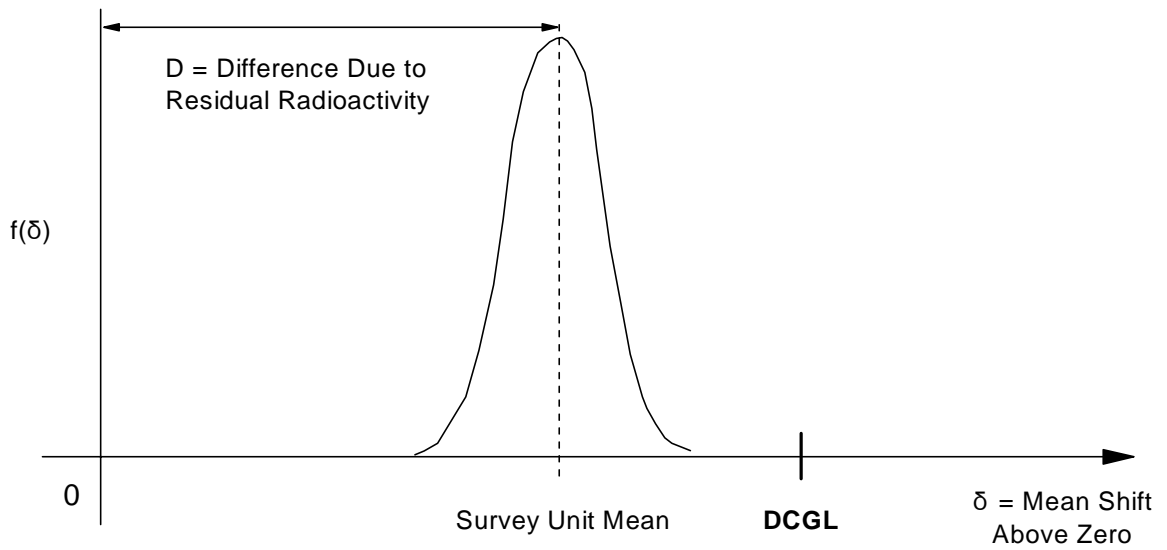
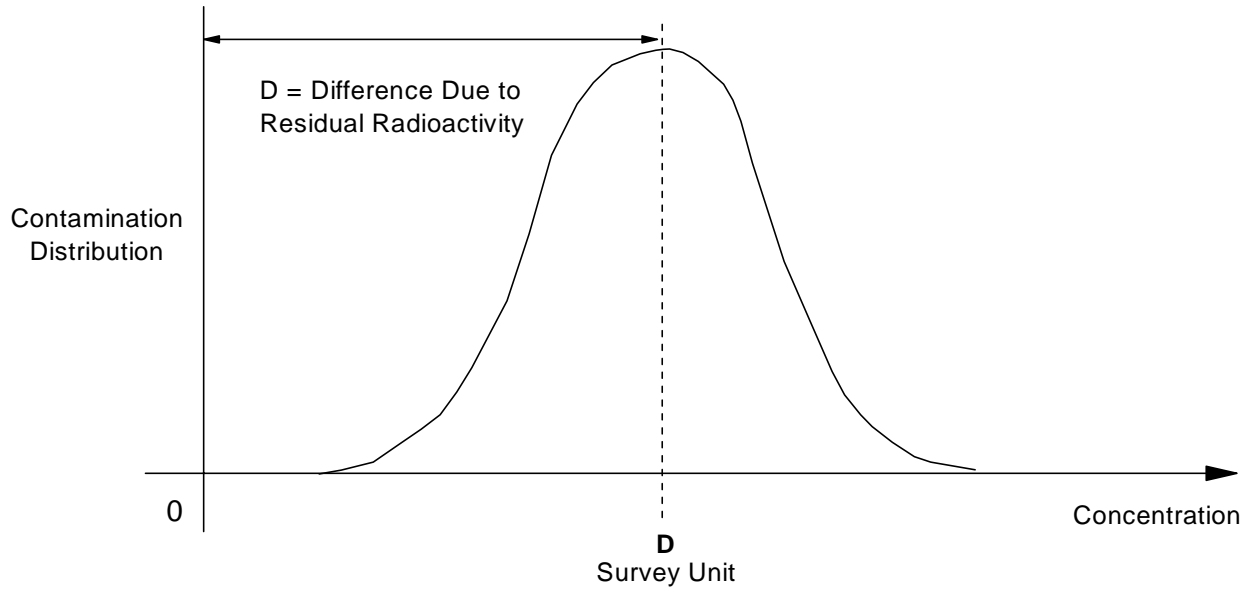
The decision rule for the 1-sample case is: “If the mean concentration in the survey unit is less than the investigation level, then the survey unit is in compliance with the release criterion.” To implement the decision rule, an estimate of the mean concentration in the survey unit is required. An estimate of the mean of the survey unit distribution may be obtained by measuring radionuclide concentrations in soil at a set of  $n$  randomly selected locations in the survey unit. A point estimate for the survey unit mean is obtained by calculating the simple arithmetic average of the  $n$  measurements. Due to measurement variability, there is a distribution of possible values for the point estimate for the survey unit mean,  $\delta$ . This distribution is referred to as  $f(\delta)$ , and is shown in the lower graph of Figure D.3. The investigation level for the Sign test used in the 1-sample case is the  $DCGL_w$ , shown on the horizontal axis of the graph.

If  $f(\delta)$  lies far to the left (or to the right) of the  $DCGL_w$ , a decision of whether or not the survey unit demonstrates compliance can be easily made. However, if  $f(\delta)$  overlaps the  $DCGL_w$ , statistical decision rules are used to assist the decision maker. Note that the width of the distribution for the estimated mean may be reduced by increasing the number of measurements. Thus, a large number of samples will reduce the probability of making decision errors.

Figure D.4 shows a simple, hypothetical example of the 2-sample case. The upper portion of the figure shows one probability distribution representing background radionuclide concentrations in the surface soil of the reference area, and another probability distribution representing radionuclide concentrations in the surface soil of the survey unit. The graph in the middle portion of the figure shows the distributions of the estimated mean concentrations in the reference area and the survey unit. In this case, the parameter of interest is the difference between the means of these two distributions,  $D$ , represented by the distance between the two vertical dotted lines.

The decision rule for the 2-sample case is: “If the difference between the mean concentration in the survey unit and the mean concentration in the reference area is less than the investigation level, then the survey unit is in compliance with the release criterion.” To implement the

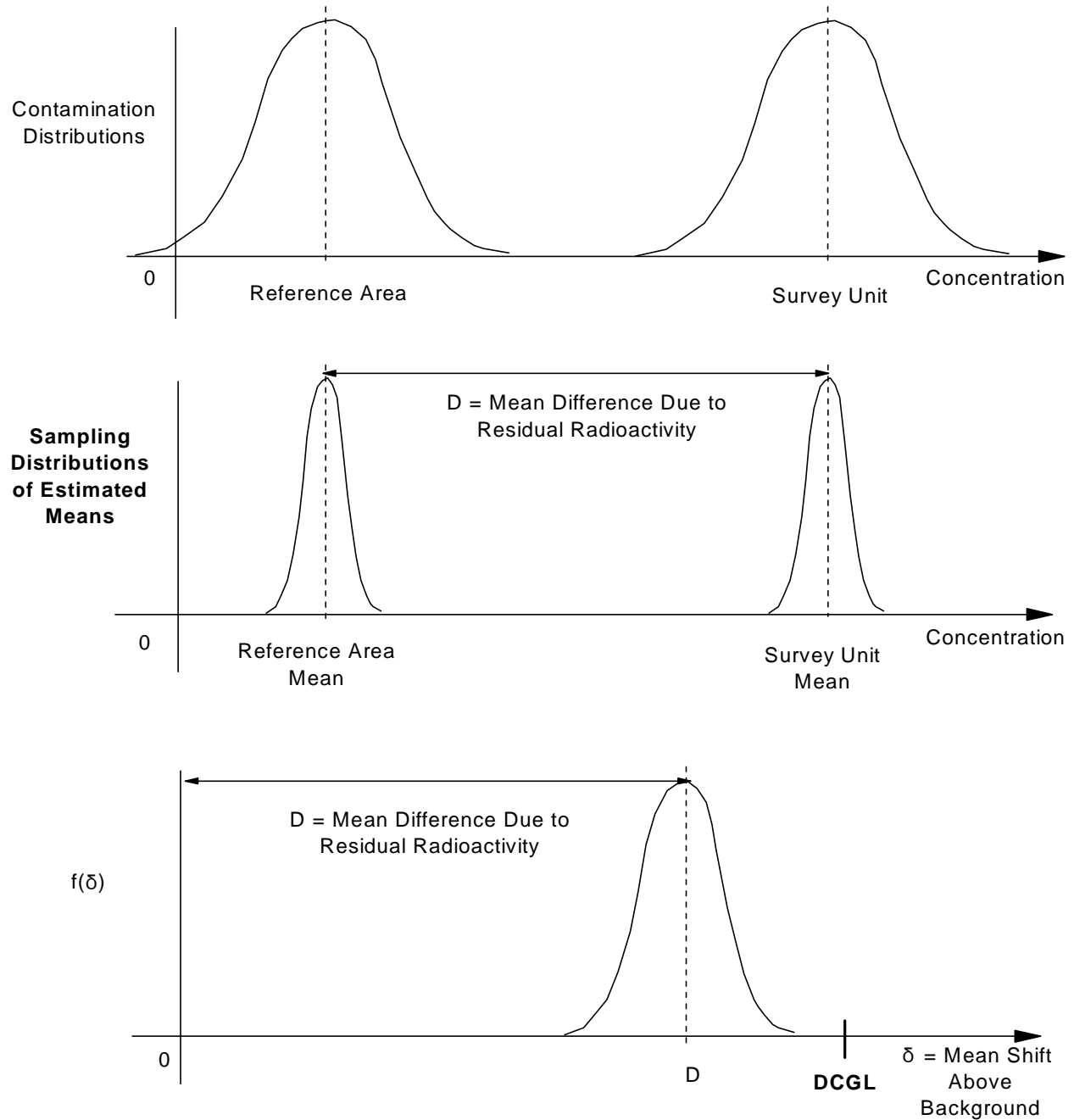
### 1-Sample Case



$f(\delta)$  is the sampling distribution of the estimated survey unit mean.

**Figure D.3 Example of the Parameter of Interest for the 1-Sample Case**

### 2-Sample Case



$f(\delta)$  is the sampling distribution of the difference between the survey unit mean and the reference area mean.

**Figure D.4 Example of the Parameter of Interest for the 2-Sample Case**

decision rule, an estimate of the difference is required. This estimate may be obtained by measuring radionuclide concentrations at a set of “n” randomly selected locations in the survey unit and “m” randomly selected locations in the reference area. A point estimate of the survey unit mean is obtained by calculating the simple arithmetic average of the n measurements in the survey unit. A point estimate of the reference area mean is similarly calculated. A point estimate of the difference between the two means is obtained by subtracting the reference area average from the survey unit average.

The measurement distribution of this difference,  $f(\delta)$ , is centered at D, the true value of the difference. This distribution is shown in the lower graph of Figure D.4.

Once again, if  $f(\delta)$  lies far to the left (or to the right) of the  $DCGL_w$ , a decision of whether or not the survey unit demonstrates compliance can be easily made. However, if  $f(\delta)$  overlaps the  $DCGL_w$ , statistical decision rules are used to assist the decision maker.

## D.6 Specify Limits on Decision Errors

Decisions based on survey results can often be reduced to a choice between “yes” or “no”, such as determining whether or not a survey unit meets the release criterion. When viewed in this way, two types of incorrect decisions, or decision errors, are identified: 1) incorrectly deciding that the answer is “yes” when the true answer is “no”, and 2) incorrectly deciding the answer is “no” when the true answer is “yes”. The distinctions between these two types of errors are important for two reasons: 1) the consequences of making one type of error versus the other may be very different, and 2) the methods for controlling these errors are different and involve tradeoffs. For these reasons, the decision maker should specify levels for each type of decision error.

The purpose of this section is to specify the decision maker's limits on decision errors, which are used to establish performance goals for the data collection design. The goal of the planning team is to develop a survey design that reduces the chance of making a decision error.

While the possibility of a decision error can never be totally eliminated, it can be controlled. To control the possibility of making decision errors, the planning team attempts to control uncertainty in the survey results caused by sampling design error and measurement error. Sampling design error may be controlled by collecting a large number of samples. Using more precise measurement techniques or field duplicate analyses can reduce measurement error. Better sampling designs can also be developed to collect data that more accurately and efficiently represent the parameter of interest. Every survey will use a slightly different method of controlling decision errors, depending on the largest source of error and the ease of reducing those error components.

## Appendix D

The estimate of the standard deviation for the measurements performed in a survey unit ( $\sigma_s$ ) includes the individual measurement uncertainty as well as the spatial and temporal variations captured by the survey design. For this reason, individual measurement uncertainties are not used during the final status survey data assessment. However, individual measurement uncertainties may be useful for determining an *a priori* estimate of  $\sigma_s$  during survey planning. Since a larger value of  $\sigma_s$  results in an increased number of measurements needed to demonstrate compliance during the final status survey, the decision maker may seek to reduce measurement uncertainty through various methods (*e.g.*, different instrumentation). There are trade-offs that should be considered during survey planning. For example, the costs associated with performing additional measurements with an inexpensive measurement system may be less than the costs associated with a measurement system with better sensitivity (*i.e.*, lower measurement uncertainty, lower minimum detectable concentration). However, the more expensive measurement system with better sensitivity may reduce  $\sigma_s$  and the number of measurements used to demonstrate compliance to the point where it is more cost effective to use the more expensive measurement system. For surveys in the early stages of the Radiation Survey and Site Investigation Process, the measurement uncertainty and instrument sensitivity become even more important. During scoping, characterization, and remedial action support surveys, decisions about classification and remediation are made based on a limited number of measurements. When the measurement uncertainty or the instrument sensitivity values approach the value of the DCGL, it becomes more difficult to make these decisions. From an operational standpoint, when operators of a measurement system have an *a priori* understanding of the sensitivity and potential measurement uncertainties, they are able to recognize and respond to conditions that may warrant further investigation—*e.g.*, changes in background radiation levels, the presence of areas of elevated activity, measurement system failure or degradation, *etc.*

The probability of making decision errors can be controlled by adopting a scientific approach, called hypothesis testing. In this approach, the survey results are used to select between one condition of the environment (the null hypothesis,  $H_0$ ) and an alternative condition (the alternative hypothesis,  $H_a$ ). The null hypothesis is treated like a baseline condition that is assumed to be true in the absence of strong evidence to the contrary. Acceptance or rejection of the null hypothesis depends upon whether or not the particular survey results are consistent with the hypothesis.

A decision error occurs when the decision maker rejects the null hypothesis when it is true, or accepts the null hypothesis when it is false. These two types of decision errors are classified as Type I and Type II decision errors, and can be represented by a table as shown in Table D.1.

A Type I decision error occurs when the null hypothesis is rejected when it is true, and is sometimes referred to as a false positive error. The probability of making a Type I decision error, or the level of significance, is denoted by alpha ( $\alpha$ ). Alpha reflects the amount of evidence the decision maker would like to see before abandoning the null hypothesis, and is also referred to as the *size* of the test.

**Table D.1 Example Representation of Decision Errors for a Final Status Survey**

		DECISION	
		Reject $H_0$ (Meets Release Criterion)	Accept $H_0$ (Exceeds Release Criterion)
TRUE CONDITION OF SURVEY UNIT	Meets Release Criterion	(No decision error)	Incorrectly Fail to Release Survey Unit (Type II)
	Exceeds Release Criterion	Incorrectly Release Survey Unit (Type I)	(No decision error)

A Type II decision error occurs when the null hypothesis is accepted when it is false. This is sometimes referred to as a false negative error. The probability of making a Type II decision error is denoted by beta ( $\beta$ ). The term  $(1-\beta)$  is the probability of rejecting the null hypothesis when it is false, and is also referred to as the *power* of the test.

There is a relationship between  $\alpha$  and  $\beta$  that is used in developing a survey design. In general, increasing  $\alpha$  decreases  $\beta$  and vice versa, holding all other variables constant. Increasing the number of measurements typically results in a decrease in both  $\alpha$  and  $\beta$ . The number of measurements that will produce the desired values of  $\alpha$  and  $\beta$  from the statistical test can be estimated from  $\alpha$ ,  $\beta$ , the  $DCGL_w$ , and the estimated variance of the distribution of the parameter of interest.

There are five activities associated with specifying limits on decision errors:

- Determining the possible range of the parameter of interest. Establish the range by estimating the likely upper and lower bounds based on professional judgement.
- Identifying the decision errors and choosing the null hypothesis.
  - a. Define both types of decision errors (Type I and Type II) and establish the true condition of the survey unit for each decision error.
  - b. Specify and evaluate the potential consequences of each decision error.
  - c. Establish which decision error has more severe consequences near the action level. Consequences include health, ecological, political, social, and resource risks.

## Appendix D

- d. Define the null hypothesis and the alternative hypothesis and assign the terms "Type I" and "Type II" to the appropriate decision error.
- Specifying a range of possible parameter values, a gray region, where the consequences of decision errors are relatively minor. It is necessary to specify a gray region because variability in the parameter of interest and unavoidable imprecision in the measurement system combine to produce variability in the data such that a decision may be "too close to call" when the true but unknown value of the parameter of interest is very near the action level. Additional guidance on specifying a gray region is available in *Guidance for the Data Quality Objectives Process* (EPA 1994a).
- Assigning probability limits to points above and below the gray region that reflect the probability for the occurrence of decision errors.
- Graphically representing the decision rule.

The expected outputs of this step are decision error rates based on the consequences of making an incorrect decision. Certain aspects of the site investigation process, such as the Historical Site Assessment (HSA), are not so quantitative that numerical values for decision errors can be specified. Nevertheless, a "comfort region" should be identified where the consequences of decision errors are relatively minor.

In Section D.5, the parameter of interest was defined as the difference between the survey unit mean concentration of residual radioactivity and the reference area mean concentration in the 2-sample case, or simply the survey unit mean concentration in the 1-sample case. The possible range of values for the parameter of interest is determined based on existing information (such as the Historical Site Assessment or previous surveys) and best professional judgement. The likely lower bound for  $f(\delta)$  is either background or zero. For a final status survey when the residual radioactivity is expected to meet the release criterion, and a conservative upper bound might be approximately three times  $DCGL_w$ .

Hypothesis testing is used to determine whether or not a statement concerning the parameter of interest should be verified. The statement about the parameter of interest is called the null hypothesis. The alternative hypothesis is the opposite of what is stated in the null hypothesis. The decision maker needs to choose between two courses of action, one associated with the null hypothesis and one associated with the alternative hypothesis.

To make a decision using hypothesis testing, a test statistic is compared to a critical value. The *test statistic*<sup>1</sup> is a number calculated using data from the survey. The critical value of the test statistic defines a rejection region based on some assumptions about the true distribution of data in the survey unit. If the value of the test statistic falls within the rejection region, the null

---

<sup>1</sup> The test statistic is not necessarily identical to the parameter of interest, but is functionally related to it through the statistical analysis.



hypothesis is rejected. The decision rule, developed in Section D.5, is used to describe the relationship between the test statistic and the critical value.

MARSSIM considers two ways to state  $H_0$  for a final status survey. The primary consideration in most situations will be compliance with the release criterion. This is shown as Scenario A in Figure D.5. The null hypothesis is that the survey unit exceeds the release criterion. Using this statement of  $H_0$  means that significant evidence that the survey unit does not exceed the release criterion is required before the survey unit would be released.

In some situations, however, the primary consideration may be determining if any residual radioactivity at the site is distinguishable from background, shown as Scenario B in Figure D.6. In this manual, Scenario A is used as an illustration because it directly addresses the compliance issue and allows consideration of decision errors. More information on Scenario B can be found in the NRC draft report NUREG-1505 (NRC 1995a).

For Scenario A, the null hypothesis is that the survey unit does not meet the release criterion. A Type I decision error would result in the release of a survey unit containing residual radioactivity above the release criterion. The probability of making this error is  $\alpha$ . Setting a high value for  $\alpha$  would result in a higher risk that survey units that might be somewhat in excess of the release criterion would be passed as meeting the release criterion. Setting a low value for  $\alpha$  would result in fewer survey units where the null hypothesis is rejected. However, the cost of setting a low value for  $\alpha$  is either a higher value for  $\beta$  or an increased number of samples used to demonstrate compliance.

For Scenario A, the alternative hypothesis is that the survey unit does meet the release criterion. A Type II decision error would result in either unnecessary costs due to remediation of survey units that are truly below the release criterion or additional survey activities to demonstrate compliance. The probability of making a Type II error is  $\beta$ . Selecting a high value for  $\beta$  (low power) would result in a higher risk that survey units that actually meet the release criterion are subject to further investigation. Selecting a low value for  $\beta$  (high power) will minimize these investigations, but the tradeoff is either a higher value for  $\alpha$  or an increased number of measurements used to demonstrate compliance. Setting acceptable values for  $\alpha$  and  $\beta$ , as well as determining an appropriate gray region, is a crucial step in the DQO process.

In the MARSSIM framework, the gray region is always bounded from above by the DCGL corresponding to the release criterion. The *Lower Bound of the Gray Region* (LBGR) is selected during the DQO process along with the target values for  $\alpha$  and  $\beta$ . The *width* of the gray region, equal to (DCGL - LBGR), is a parameter that is central to the nonparametric tests discussed in this manual. It is also referred to as the *shift*,  $\Delta$ . The absolute size of the shift is actually of less importance than the *relative shift*  $\Delta/\sigma$ , where  $\sigma$  is an estimate of the standard deviation of the measured values in the survey unit. The estimated standard deviation,  $\sigma$ , includes both the real spatial variability in the quantity being measured, and the precision of the chosen measurement

**SCENARIO A**

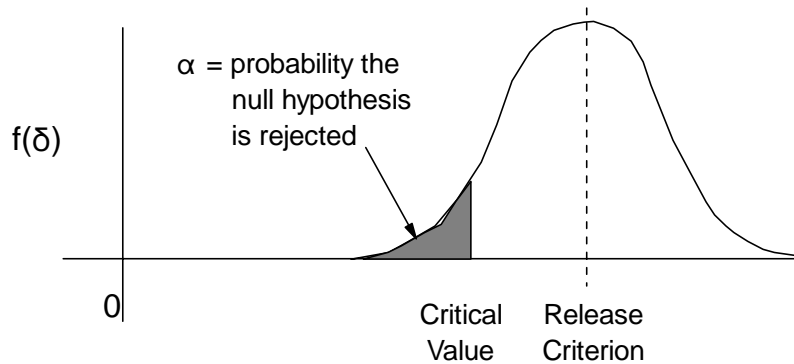
Assume as a null hypothesis that the survey unit exceeds the release criterion. This requires significant evidence that the residual radioactivity in the survey unit is less than the release criterion to reject the null hypothesis (and pass the survey unit). If the evidence is not significant at level  $\alpha$ , the null hypothesis of a non-complying survey unit is accepted (and the survey unit fails).

**HYPOTHESIS TEST**

$H_0$ : Survey unit does not meet release criterion

$H_a$ : Survey unit does meet the release criterion

Survey unit passes if and only if the test statistic falls in the rejection region.



This test directly addresses the compliance question.

The mean shift for the survey unit must be significantly below the release criterion for the null hypothesis to be rejected.

With this test, site owners face a trade-off between additional sampling costs and unnecessary remediation costs. They may choose to increase the number of measurements in order to decrease the number of Type II decision errors (reduce the chance of remediating a clean survey unit for survey units at or near background levels).

Distinguishability from background is not directly addressed. However, sample sizes may be selected to provide adequate power at or near background levels, hence ensuring that most survey units near background would pass. Additional analyses, such as point estimates and/or confidence intervals, may be used to address this question.

A high percentage of survey units slightly below the release criterion may fail the release criterion, unless large numbers of measurements are used. This achieves a high degree of assurance that most survey units that are at or above the release criterion will not be improperly released.

**Figure D.5 Possible Statement of the Null Hypothesis for the Final Status Survey Addressing the Issue of Compliance**

### SCENARIO B

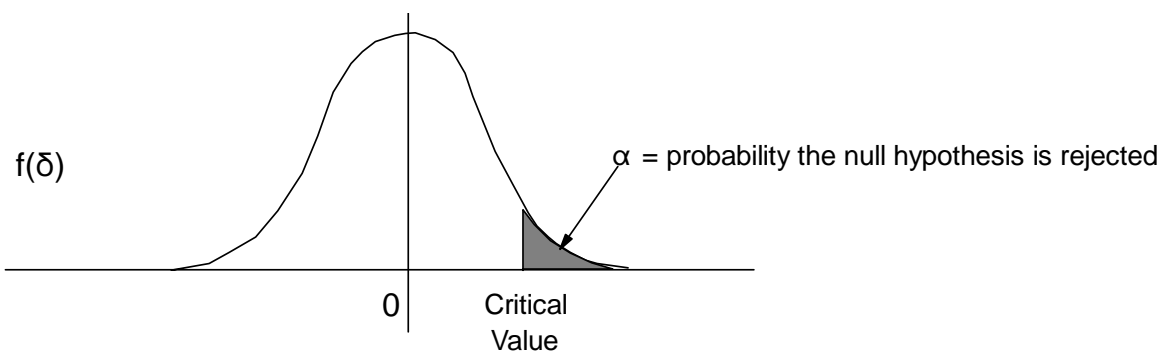
Assume as a null hypothesis that the survey unit is indistinguishable from background. This requires significant evidence that the survey unit residual radioactivity is greater than background to reject the null hypothesis (and fail the survey unit). If the evidence is not significant at level  $\alpha$ , the null hypothesis of a clean survey unit is accepted (and the survey unit passes).

#### HYPOTHESIS TEST

$H_0$ : Survey unit is indistinguishable from background

$H_a$ : Survey unit is distinguishable from background

Survey unit passes if and only if the test statistic falls in the rejection region.



Distinguishability from background may be of primary importance to some stakeholders.

The residual radioactivity in the survey unit must be significantly above background for the null hypothesis to be rejected.

Compliance with the DCGLs is not directly addressed. However, the number of measurements may be selected to provide adequate power at or near the DCGL, hence ensuring that most survey units near the DCGL would not be improperly released. Additional analysis, based on point estimates and/or confidence intervals, is required to determine compliance if the null hypothesis is rejected by the test.

A high percentage of survey units slightly below the release criterion will fail unless large numbers of measurements are used. This is necessary to achieve a high degree of assurance that for most sites at or above the release criterion the null hypothesis will fail to be improperly released.

**Figure D.6 Possible Statement of the Null Hypothesis for the Final Status Survey Addressing the Issue of Indistinguishability from Background**

method. The relative shift,  $\Delta/\sigma$ , is an expression of the resolution of the measurements in units of measurement uncertainty. Expressed in this way, it is easy to see that relative shifts of less than one standard deviation,  $\Delta/\sigma < 1$ , will be difficult to detect. On the other hand, relative shifts of more than three standard deviations,  $\Delta/\sigma > 3$ , are generally easier to detect. The number of measurements that will be required to achieve given error rates,  $\alpha$  and  $\beta$ , depends almost entirely on the value of  $\Delta/\sigma$  (see Chapter 5).

Since small values of  $\Delta/\sigma$  result in large numbers of samples, it is important to design for  $\Delta/\sigma > 1$  whenever possible. There are two obvious ways to increase  $\Delta/\sigma$ . The first is to increase the width of the gray region by making LBGR small. Only Type II decision errors occur in the gray region. The disadvantage of making this gray region larger is that the probability of incorrectly failing to release a survey unit will increase. The target false negative rate  $\beta$  will be specified at lower residual radioactivity levels, i.e., a survey unit will generally have to be lower in residual radioactivity to have a high probability of being judged to meet the release criterion. The second way to increase  $\Delta/\sigma$  is to make  $\sigma$  smaller. One way to make  $\sigma$  small is by having survey units that are relatively homogeneous in the amount of measured radioactivity. This is an important consideration in selecting survey units that have both relatively uniform levels of residual radioactivity and also have relatively uniform background radiation levels. Another way to make  $\sigma$  small is by using more precise measurement methods. The more precise methods might be more expensive, but this may be compensated for by the decrease in the number of required measurements. One example would be in using a radionuclide specific method rather than gross radioactivity measurements for residual radioactivity that does not appear in background. This would eliminate the variability in background from  $\sigma$ , and would also eliminate the need for reference area measurements.

The effect of changing the width of the gray region and/or changing the measurement variability on the estimated number of measurements (and cost) can be investigated using the DEFT (Decision Error Feasibility Trials) software developed by EPA (EPA 1995a). This program can only give approximate sample sizes and costs since it assumes that the measurement data are normally distributed, that a Student's t test will be used to evaluate the data, and that there is currently no provision for comparison to a reference area. Nevertheless, as a rough rule of thumb, the sample sizes calculated by DEFT are about 85% of those required by the one-sample nonparametric tests recommended in this manual. This rule of thumb works better for large numbers of measurements than for smaller numbers of measurements, but can be very useful for estimating the relative impact on costs of decisions made during the planning process.

Generally, the design goal should be to achieve  $\Delta/\sigma$  values between one and three. The number of samples needed rises dramatically when  $\Delta/\sigma$  is smaller than one. Conversely, little is usually gained by making  $\Delta/\sigma$  larger than about three. If  $\Delta/\sigma$  is greater than three or four, one should take advantage of the measurement precision available by making the width of the gray region smaller. It is even more important, however, that overly optimistic estimates for  $\sigma$  be avoided. The consequence of taking fewer samples than are needed given the actual measurement variations will be unnecessary remediations (increased Type II decision errors).

Once the preliminary estimates of  $\Delta$  and  $\sigma$  are available, target values for  $\alpha$  and  $\beta$  can be selected. The values of  $\alpha$  and  $\beta$  should reflect the risks involved in making Type I and Type II decision errors, respectively.

One consideration in setting the false positive rate are the health risks associated with releasing a survey unit that might actually contain residual radioactivity in excess of the  $DCGL_w$ . If a survey unit did exceed the  $DCGL_w$ , the first question that arises is “How much above the  $DCGL_w$  is the residual radioactivity likely to be?” The DEFT software can be used to evaluate this.

For example, if the  $DCGL_w$  is 100 Bq/kg (2.7 pCi/g), the LBGR is 50 Bq/kg (1.4 pCi/g),  $\sigma$  is 50 Bq/kg (1.4 pCi/g),  $\alpha = 0.10$  and  $\beta = 0.05$ , the DEFT calculations show that while a survey unit with residual radioactivity equal to the  $DCGL_w$  has a 10% chance of being released, a survey unit at a level of 115 Bq/kg (3.1 pCi/g) has less than a 5% chance of being released, a survey unit at a level of 165 Bq/kg (4.5 pCi/g) has virtually no chance of being released. However, a survey unit with a residual radioactivity level of 65 Bq/kg (1.8 pCi/g) will have about an 80% chance of being released and a survey unit with a residual radioactivity level of 80 Bq/kg (2.2 pCi/g) will only have about a 40% chance of being released. Therefore, it is important to examine the probability of deciding that the survey unit does not meet the release criterion over the entire range of possible residual radioactivity values, and not only at the boundaries of the gray region. Of course, the gray region can be made narrower, but at the cost of additional sampling. Since the equations governing the process are not linear, small changes can lead to substantial changes in survey costs.

As stated earlier, the values of  $\alpha$  and  $\beta$  that are selected in the DQO process should reflect the risk involved in making a decision error. In setting values for  $\alpha$ , the following are important considerations:

- In radiation protection practice, public health risk is modeled as a linear function of dose (BEIR 1990). Therefore a 10% change in dose, say from 15 to 16.5, results in a 10% change in risk. This situation is quite different from one in which there is a threshold. In the latter case, the risk associated with a decision error can be quite high, and low values of  $\alpha$  should be selected. When the risk is linear, much higher values of  $\alpha$  at the release criterion might be considered adequately protective when the survey design results in smaller decision error rates at doses or risks greater than the release criterion. False positives will tend to be balanced by false negatives across sites and survey units, resulting in approximately equal human health risks.
- The DCGL itself is not free of error. The dose or risk cannot be measured directly, and many assumptions are made in converting doses or risks to derived concentrations. To be adequately protective of public health, these models are generally designed to over predict the dose or risk. Unfortunately, it is difficult to quantify this. Nonetheless, it is probably safe to say that most models have uncertainty sufficiently large such that the true dose or risk delivered by residual radioactivity at the DCGL is very likely to be lower than the

release criterion. This is an additional consideration for setting the value of  $\alpha$ , that could support the use of larger values in some situations. In this case, one would prospectively address, as part of the DQO process, the magnitude, significance, and potential consequences of decision errors at values above the release criterion. The assumptions made in any model used to predict DCGLs for a site should be examined carefully to determine if the use of site specific parameters results in large changes in the DCGLs, or whether a site-specific model should be developed rather than designing a survey around DCGLs that may be too conservative.

- The risk of making the second type of decision error,  $\beta$ , is the risk of requiring additional remediation when a survey unit already meets the release criterion. Unlike the health risk, the cost associated with this type of error may be highly non-linear. The costs will depend on whether the survey unit has already had remediation work performed on it, and the type of residual radioactivity present. There may be a threshold below which the remediation cost rises very rapidly. If so, a low value for  $\beta$  is appropriate at that threshold value. This is primarily an issue for survey units that have a substantial likelihood of falling at or above the gray region for residual radioactivity. For survey units that are very lightly contaminated, or have been so thoroughly remediated that any residual radioactivity is expected to be far below the DCGL, larger values of  $\beta$  may be appropriate especially if final status survey sampling costs are a concern. Again, it is important to examine the probability of deciding that the survey unit does not meet the release criterion over the entire range of possible residual radioactivity values, below as well as above the gray region.
- Lower decision error rates may be possible if alternative sampling and analysis techniques can be used that result in higher precision. The same might be achieved with moderate increases in sample sizes. These alternatives should be explored before accepting higher design error rates. However, in some circumstances, such as high background variations, lack of a radionuclide specific technique, and/or radionuclides that are very difficult and expensive to quantify, error rates that are lower than the uncertainties in the dose or risk estimates may be neither cost effective nor necessary for adequate radiation protection.

None of the above discussion is meant to suggest that under any circumstances a less than rigorous, thorough, and professional approach to final status surveys would be satisfactory. The decisions made and the rationale for making these decisions should be thoroughly documented.

For Class 1 Survey Units, the number of samples may be driven more by the need to detect small areas of elevated activity than by the requirements of the statistical tests. This in turn will depend primarily on the sensitivity of available scanning instrumentation, the size of the area of elevated activity, and the dose or risk model. A given concentration of residual radioactivity spread over a smaller area will, in general, result in a smaller dose or risk. Thus, the  $DCGL_{EMC}$  used for the elevated measurement comparison is usually larger than the  $DCGL_w$  used for the statistical test. In some cases, especially radionuclides that deliver dose or risk primarily via internal pathways,

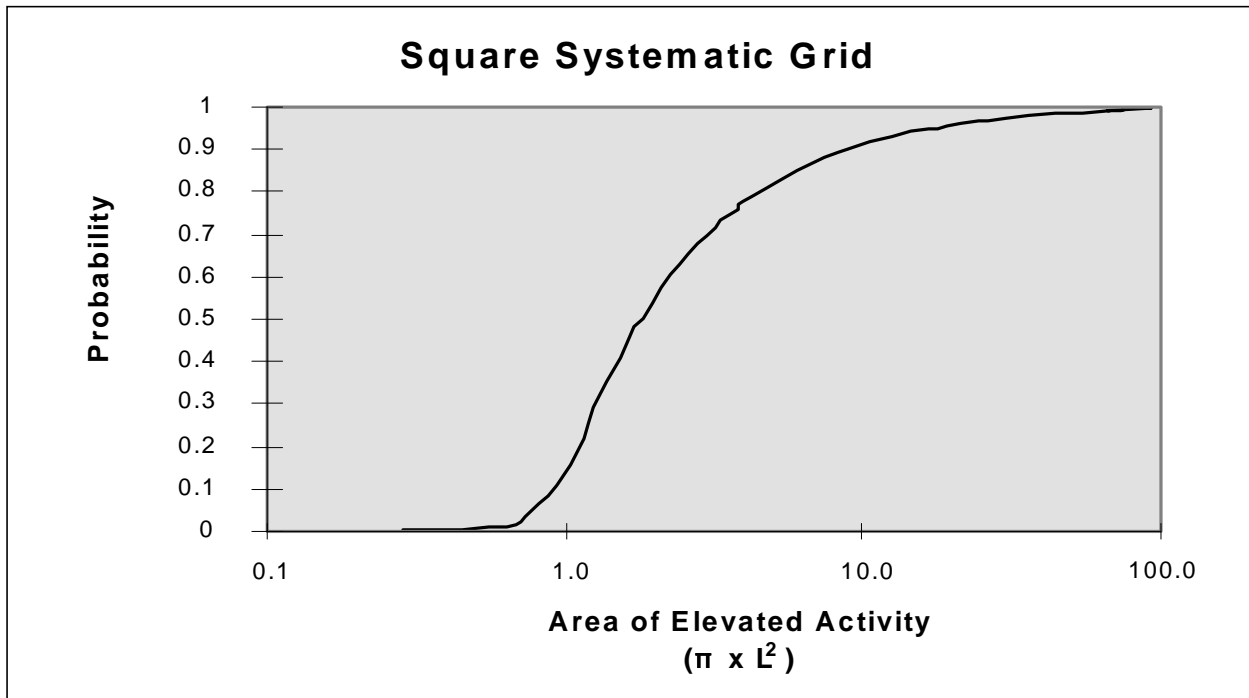
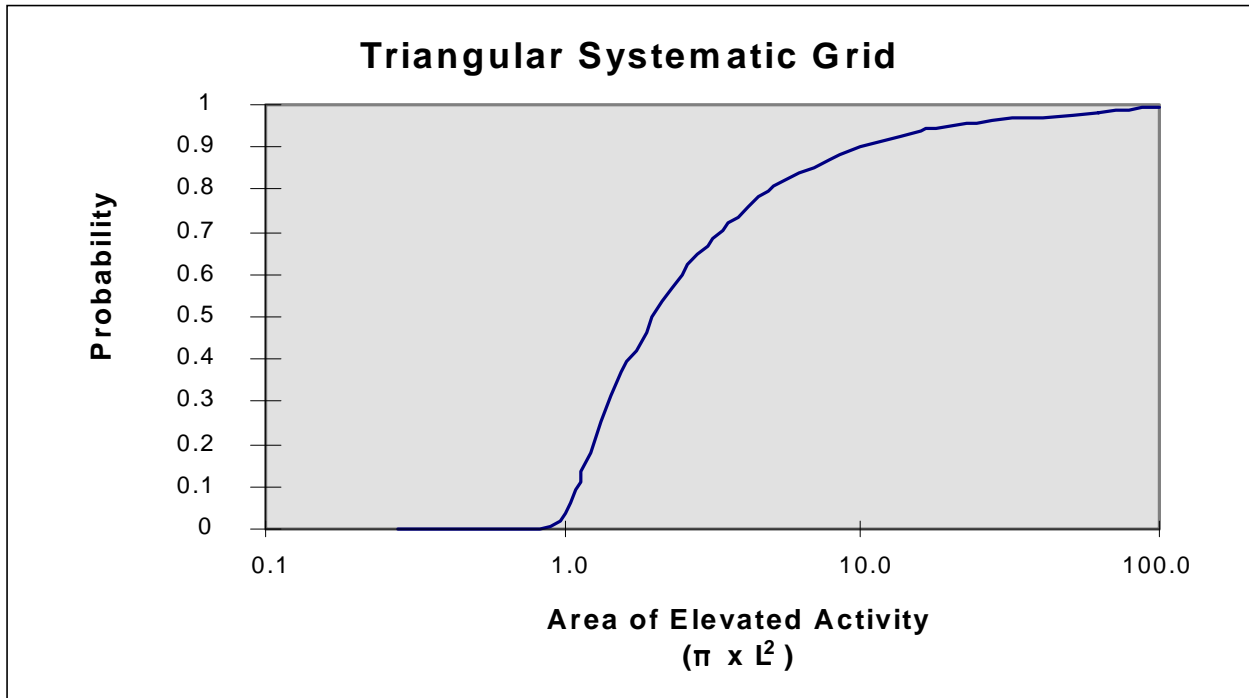
dose or risk is approximately proportional to inventory, and so the difference in the DCGLs is approximately proportional to the areas.

However, this may not be the case for radionuclides that deliver a significant portion of the dose or risk via external exposure. The exact relationship between the  $DCGL_{EMC}$  and the  $DCGL_w$  is a complicated function of the dose or risk modeling pathways, but area factors to relate the two DCGLs can be tabulated for most radionuclides (see Chapter 5), and site-specific area factors can also be developed.

For many radionuclides, scanning instrumentation is readily available that is sensitive enough to detect residual radioactivity concentrations at the  $DCGL_{EMC}$  derived for the sampling grid of direct measurements used in the statistical tests. Where instrumentation of sufficient sensitivity (MDC, see Chapter 6) is not available, the number of samples in the survey unit can be increased until the area between sampling points is small enough (and the resulting area factor is large enough) that  $DCGL_{EMC}$  can be detected by scanning. The details of this process are discussed in Chapter 5. For some radionuclides (*e.g.*,  $^3H$ ) the scanning sensitivity is so low that this process would never terminate—*i.e.*, the number of samples required could increase without limit. Thus, an important part of the DQO process is to determine the smallest size of an area of elevated activity that it is important to detect,  $A_{min}$ , and an acceptable level of risk,  $R_A$ , that it may go undetected. The probability of sampling a circular area of size  $A$  with either a square or triangular sampling pattern is shown in Figure D.7. The ELIPGRID-PC (Davidson 1995) computer code can also be used to calculate these probabilities.

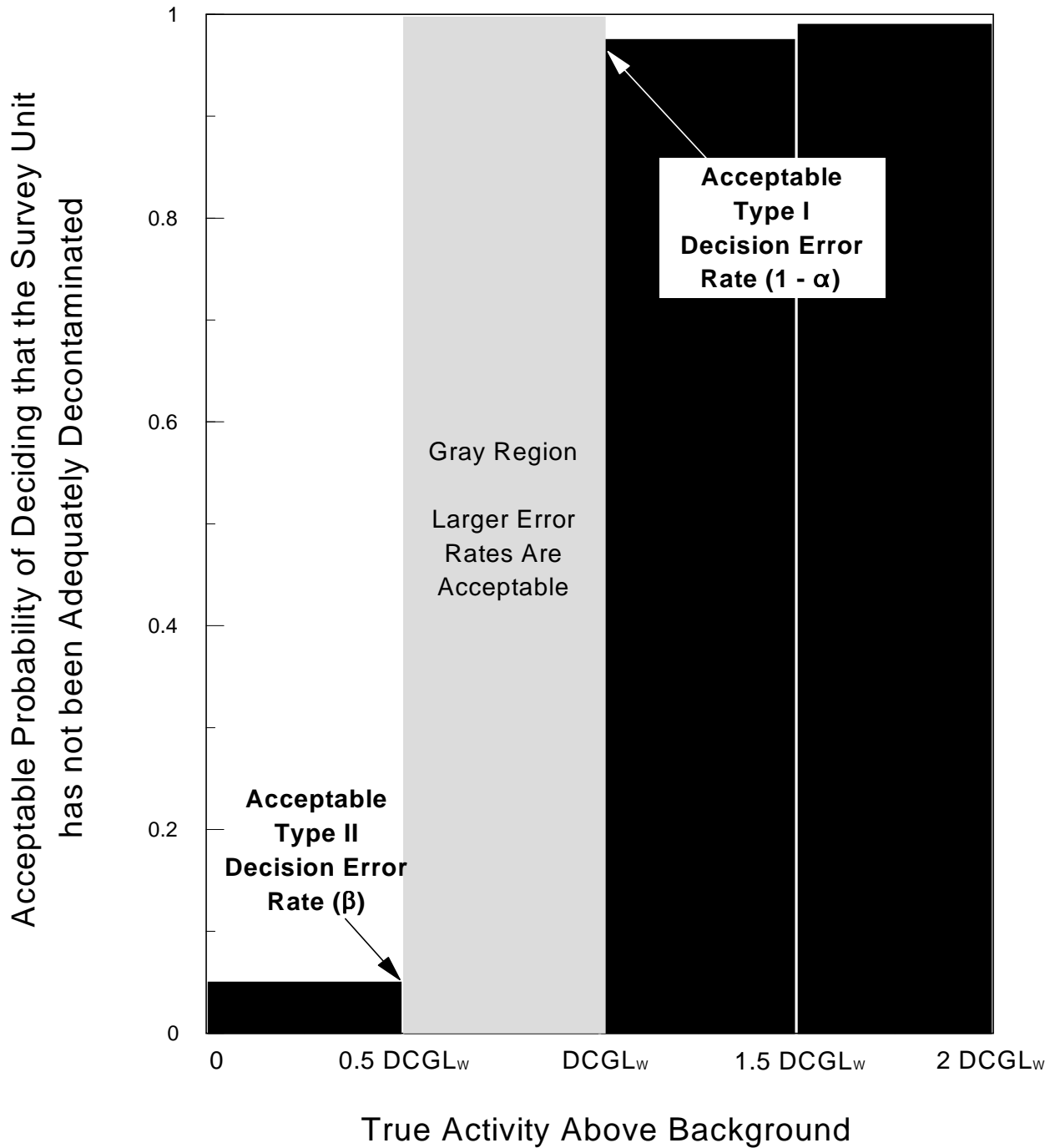
In this part of the DQO process, the concern is less with areas of elevated activity that are found than with providing adequate assurance that negative scanning results truly demonstrate the absence of such areas. In selecting acceptable values for  $A_{min}$  and  $R_A$ , maximum use of information from the HSA and all surveys prior to the final status surveys should be used to determine what sort of areas of elevated activity could possibly exist, their potential size and shape, and how likely they are to exist. When the detection limit of the scanning technique is very large relative to the  $DCGL_{EMC}$ , the number of measurements estimated to demonstrate compliance using the statistical tests may become unreasonably large. In this situation an evaluation of the survey objectives and considerations be performed. These considerations may include the survey design and measurement methodology, exposure pathway modeling assumptions and parameter values used to determine the DCGLs, Historical Site Assessment conclusions concerning source terms and radionuclide distributions, and the results of scoping and characterization surveys. In most cases the results of this evaluation is not expected to justify an unreasonably large number of measurements.

A convenient method for visualizing the decision rule is to graph the probability of deciding that the survey unit does not meet the release criterion, *i.e.*, that the null hypothesis of Scenario A is accepted. An example of such a chart is shown in Figure D.8.



**Figure D.7 Geometric Probability of Sampling at Least One Point of an Area of Elevated Activity as a Function of Sample Density with Either a Square or Triangular Sampling Pattern**





**Figure D.8 Example of a Power Chart Illustrating the Decision Rule for the Final Status Survey**

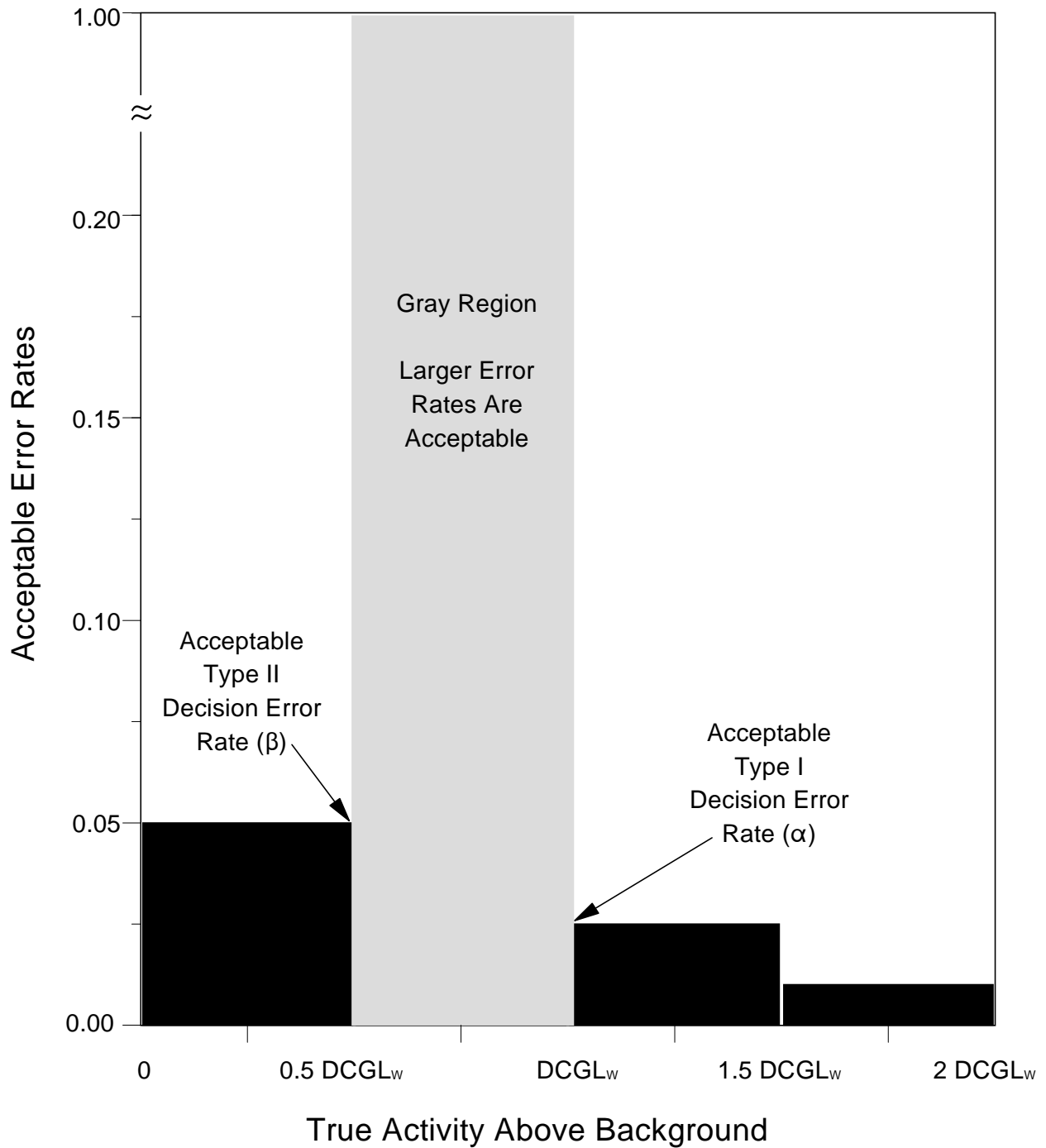
## Appendix D

In this example  $\alpha$  is 0.025 and  $\beta$  is 0.05, providing an expected power ( $1-\beta$ ) of 0.95 for the test. A second method for presenting the information is shown in Figure D.9. This figure shows the probability of making a decision error for possible values of the parameter of interest, and is referred to as an error chart. In both examples a gray region, where the consequences of decision errors are deemed to be relatively minor, is shown. These charts are used in the final step of the DQO Process, combined with the outputs from the previous steps, to produce an efficient and cost-effective survey design. It is clear that setting acceptable values for  $\alpha$  and  $\beta$ , as well as determining an appropriate gray region, is a crucial step in the DQO Process. Instructions for creating a prospective power curve, which can also be used to visualize the decision rule, are provided in Appendix I.

After the survey design is implemented, the expected values of  $\alpha$  and  $\beta$  determined in this step are compared to the actual significance level and power of the statistical test based on the measurement results during the assessment phase of the Data Life Cycle. This comparison is used to verify that the objectives of the survey have been achieved.

EPA QA/G-9 (EPA 1996a) discusses considerations for selecting a particular null hypothesis. Because of the basic hypothesis testing philosophy, the null hypothesis is generally specified in terms of the *status quo* (e.g., no change or action will take place if the null hypothesis is not rejected). Also, since the classical hypothesis testing approach exercises direct control over the Type I (false positive) error rate, this rate is generally associated with the error of most concern. In the case of the null hypothesis in which the residual radioactivity in the survey unit exceeds the release criterion, a Type I decision error would conclude that the residual activity was less than the release criterion when in fact it was above the release criterion. One difficulty, therefore, may be obtaining a consensus on which error should be of most concern (i.e., releasing a site where the residual activity exceeds the release criterion or failing to release a site where the residual activity is less than the release criterion). It is likely that the regulatory agency's public health-based protection viewpoint will differ from the viewpoint of the regulated party. The ideal approach is not only to define the null hypothesis in such a way that the Type I decision error protects human health and the environment but also in a way that encourages quality (high precision and accuracy) and minimizes expenditure of resources in situations where decisions are relatively "easy" (e.g., all observations are far below the threshold level of interest or DCGL).

To avoid excessive expense in performing measurements, compromises are sometimes necessary. For example, suppose that a significance level ( $\alpha$ ) of 0.05 is to be used. However, the affordable sample size may be expected to yield a test with power ( $\beta$ ) of only 0.40 at some specified parameter value chosen to have practical significance. One possible compromise may be to relax the Type I decision error rate ( $\alpha$ ) and use a value of 0.10, 0.15, or even 0.20. By relaxing the Type I decision error rate, a higher power (i.e., a lower Type II decision error rate) can be achieved. An argument can be made that survey designs should be developed and number of measurements determined in such a way that both the Type I ( $\alpha$ ) and Type II ( $\beta$ ) decision error rates are treated simultaneously and in a balanced manner (i.e.,  $\alpha = \beta = 0.15$ ). This approach of



**Figure D.9 Example of an Error Chart Illustrating the Decision Rule for the Final Status Survey**

treating the Type I and Type II decision error rates simultaneously is taken by the DQO Process. It is recommended that several different values for  $\alpha$  and  $\beta$  be investigated before specific values are selected.

## D.7 Optimize the Design for Collecting Data

This step is designed to produce the most resource-effective survey design that is expected to meet the DQOs. It may be necessary to work through this step more than once after revisiting previous steps in the DQO Process.

There are six activities included in this step:

- Reviewing the DQO outputs and existing environmental data to ensure they are internally consistent.
- Developing general data collection design alternatives. Chapter 5 describes random and systematic sampling designs recommended for final status surveys based on survey unit classification.
- Formulating the mathematical expressions needed to solve the design problem for each data collection design alternative.
- Selecting the optimal design that satisfies the DQOs for each data collection design alternative. If the recommended design will not meet the limits on decision errors within the budget or other constraints, then the planning team will need to relax one or more constraints. Examples include:
  - a. increasing the budget for sampling and analysis
  - b. using exposure pathway modeling to develop site-specific DCGLs
  - c. increasing the decision error rates, not forgetting to consider the risks associated with making an incorrect decision
  - d. increasing the width of the gray region by decreasing the LBGR
  - e. relaxing other project constraints—*e.g.*, schedule
  - f. changing the boundaries—it may be possible to reduce measurement costs by changing or eliminating survey units that will require different decisions
  - g. evaluating alternative measurement techniques with lower detection limits or lower survey costs
  - h. considering the use of passive controls when releasing the survey unit rather than unrestricted release
- Selecting the most resource-effective survey design that satisfies all of the DQOs. Generally, the survey designs described in Chapter 5 will be acceptable for demonstrating compliance. Atypical sites (*e.g.*, mixed-waste sites) may require the planning team to consider alternative survey designs on a site-specific basis.

- Documenting the operational details and theoretical assumptions of the selected design in the QAPP, the field sampling plan, the sampling and analysis plan, or the decommissioning plan. All of the decisions that will be made based on the data collected during the survey should be specified along with the alternative actions that may be adopted based on the survey results.

Chapters 4 and 5 present a framework for a final status survey design. When this framework is combined with the site-specific DQOs developed using the guidance in this section, the survey design should be acceptable for most sites. The key inputs to Chapters 4 and 5 are:

- investigation levels and DCGLs for each radionuclide of interest
- acceptable measurement techniques for scanning, sampling, and direct measurements, including detection limits and estimated survey costs
- identification and classification of survey units
- an estimate of the variability in the distribution of residual radioactivity for each survey unit, and in the reference area if necessary
- the decision maker's acceptable *a priori* values for decision error rates ( $\alpha$  and  $\beta$ )