APPENDIX N

Data Validation Using Data Descriptors

Data validation is often defined by six data descriptors:

1) reports to decision maker
2) documentation
3) data sources
4) analytical method and detection limit
5) data review
6) data quality indicators

The decision maker or reviewer examines the data, documentation, and reports for each of the six data descriptors to determine if performance is within the limits specified in the DQOs developed during survey planning. The data validation process should be conducted according to procedures documented in the QAPP.

N.1 Reports to Decision Maker

Data and documentation supplied to the decision maker should be evaluated for completeness and appropriateness and to determine if any changes were made to the survey plan during the course of work. The survey plan discusses the surveying, sampling, and analytical design and contains the QAPP and DQOs. The decision maker should receive all data as collected plus preliminary and final data reports. The final decision on qualifying or rejecting data will be made during the assessment of environmental data. All data, including qualified or rejected data, should be documented and recorded even if the data are not included in the final report.

Preliminary analytical data reports allow the decision maker to begin the assessment process as soon as the surveying effort has begun. These initial reports have three functions.

1) For scoping or characterization survey data, they allow the decision maker to begin to characterize the site on the basis of actual data. Radionuclides of interest will be identified and the variability in concentration can be estimated.

2) They allow potential measurement problems to be identified and the need for corrective action can be assessed.

3) Schedules are more likely to be met if the planning of subsequent survey activities can begin before the final data reports are produced.

N.2 Documentation

Three types of documentation should be assessed: (1) field operation records; (2) laboratory records; and (3) data handling records (EPA 1997a).

N.2.1 Field Operation Records

The information contained in these records documents overall field operations and generally consists of the following:

- *Field measurement records*. These records show that the proper measurement protocol was performed in the field. At a minimum, this documentation should include the names of the persons conducting the activity, measurement identification, measurement locations, measurement results, maps and diagrams, equipment and SOP used, and unusual observations. Bound field notebooks are generally used to record raw data and make references to prescribed procedures and changes in planned activities. Data recording forms might also be used. A document control system should be used for these records to control attributes such as formatting to include pre-numbered pages with date and signature lines.

- *Sample tracking records*. Sample tracking records (*e.g.*, chain-of-custody) document the progression of samples as they travel from the original sampling location to the laboratory and finally to disposal (see Section 7.7).

- *QC measurement records*. QC measurement records document the performance of QC measurements in the field. These records should include calibration and standards' traceability documentation that can be used to provide a reproducible reference point to which all similar measurements can be correlated. QC measurement records should contain information on the frequency, conditions, level of standards, and instrument calibration history.

- *Personnel files*. Personnel files record the names and training certificates of the staff collecting the data.

- *General field procedures*. General field procedures (*e.g.*, SOPs) record the procedures used in the field to collect data and outline potential areas of difficulty in performing measurements.

- *Deficiency and problem identification reports*. These reports document problems and deficiencies encountered as well as suggestions for process improvement.

- *Corrective action reports*. Corrective action reports show what methods were used in cases where general field practices or other standard procedures were violated and include the methods used to resolve noncompliance.

N.2.2 Laboratory Records

The following list describes some of the laboratory-specific records that should be compiled if available and appropriate:

- *Laboratory measurement results and sample data.* These records contain information on the sample analysis used to verify that prescribed analytical methods were followed. The overall number of samples, sample identification, sample measurement results, any deviations from the SOPs, time of day, and date should be included. Sample location information might also be provided.

- *Sample management records*. Sample management records should document sample receipt, handling and storage, and scheduling of analyses. The records will verify that sample tracking requirements were maintained, reflect any anomalies in the samples (*e.g.*, receipt of damaged samples), and note proper log-in of samples into the laboratory.

- *Test methods*. Unless analyses were performed exactly as prescribed by SOPs, this documentation will describe how the analyses were carried out in the laboratory. This documentation includes sample preparation and analysis, instrument standardization, detection and reporting limits, and method-specific QC requirements. Documentation demonstrating laboratory proficiency with each method used could also be a part of the data reporting package, particularly for subcontracted work.

- *QC measurement records*. These include the general QC records, such as initial demonstration of capability, instrument calibration, routine monitoring of analytical performance, calibration verification, *etc.*, considered in Section 7.3 for selecting a radioanalytical laboratory. Project-specific information from the QC checks such as blanks, spikes, calibration check samples, replicates, splits, and so on should be included in these reports to facilitate data quality analysis.

- *Deficiency and problem identification reports*. These reports document problems and deficiencies encountered as well as suggestions for process improvement.

- *Corrective action reports*. Corrective action reports show what methods were used in cases where general laboratory practices or other standard procedures were violated and include the methods used to resolve noncompliance. Corrective action procedures to replace samples violating the SOP also should be noted.

N.2.3 Data Handling Records

Data handling records document protocols used in data reduction, verification, and validation. Data reduction addresses data transformation operations such as converting raw data into reportable quantities and units, using significant figures, calculating measurement uncertainties, *etc.* The records document procedures for handling data corrections.

N.3 Data Sources

Data source assessment involves the evaluation and use of historical analytical data. Historical analytical data should be evaluated according to data quality indicators and not the source of the data (*e.g.*, analytical protocols may have changed significantly over time). Data quality indicators are qualitative and quantitative descriptors used in interpreting the degree of acceptability or utility of data. Historical data sources are addressed during the Historical Site Assessment, and are discussed in Section 3.4.1.

N.4 Analytical Method and Detection Limit

The selection of appropriate analytical methods based on detection limits is important to survey planning. The detection limit of the method directly affects the usability of the data because results near the detection limit have a greater possibility of false negatives and false positives. Results near the detection limit have increased measurement uncertainty. When the measurement uncertainty becomes large compared to the variability in the radionuclide concentration, it becomes more difficult to demonstrate compliance using the guidance provided in MARSSIM.

The decision maker compares detection limits (*i.e.*, minimum detectable concentrations; MDCs) with radionuclide-specific results to determine their effectiveness in relation to the DCGL. Assessment of preliminary data reports provides an opportunity to review the detection limits early and resolve any detection sensitivity problems. When a radionuclide is reported as not detected, the result can only be used with confidence if the MDCs reported are lower than the DCGL.

If the DCGL is less than or equal to the MDC, and the radionuclide is not detected, report the actual result of the analysis. Do not report data as “less than the detection limit.” Even negative results and results with large uncertainties can be used in the statistical tests described in Chapter 8. Results reported as “<MDC” cannot be fully used and, for example, complicate even such simple analyses as calculating an average. When the MDC reported for a radionuclide is near the DCGL, the confidence in both identification and quantitation may be low. Information

concerning non-detects or detections at or near MDCs should be qualified according to the degree of acceptable uncertainty.

N.5 Data Review

Data review begins with an assessment of the quality of analytical results and is performed by a professional with knowledge of the analytical procedures. Only data that are reviewed according to a specified level or plan should be used in the quantitative site investigation. Any analytical errors, or limitations in the data that are identified by the review, should be noted. An explanation of data qualifiers should be included with the review report.

All data should receive some level of review. Data that have not been reviewed should be identified, because the lack of review increases the uncertainty in the data. Unreviewed data may lead to Type I and Type II decision errors, and may also contain transcription errors and calculation errors. Data may be used in the preliminary assessment before review, but should be reviewed at a predetermined level before use in the final survey report.

Depending on the survey objectives, the level and depth of the data review varies. The level and depth of the data review may be determined during the planning process and should include an examination of laboratory and method performance for the measurements and radionuclides involved. This examination includes

- evaluation of data completeness
- verification of instrument calibration
- measurement of precision using duplicates, replicates, or split samples
- measurement of bias using reference materials or spikes
- examination of blanks for contamination
- assessment of adherence to method specifications and QC limits
- evaluation of method performance in the sample matrix
- applicability and validation of analytical procedures for site-specific measurements
- assessment of external QC measurement results and QA assessments

A different level or depth of data review may be indicated by the results of this evaluation. Specific data review procedures are dependent upon the survey objectives and should be documented in the QAPP.

N.6 Data Quality Indicators

The assessment of data quality indicators presented in this section is significant to determine data usability. The principal data quality indicators are precision, bias, representativeness, comparability, and completeness (EPA 1997a). Other data quality indicators affecting the RSSI process include the selection and classification of survey units, Type I and Type II decision error rates, the variability in the radionuclide concentration measured within the survey unit, and the lower bound of the gray region (see Section 2.3.1).

Of the six principal data quality indicators, precision and bias are quantitative measures, representativeness and comparability are qualitative, completeness is a combination of both qualitative and quantitative measures, and accuracy is a combination of precision and bias. The selection and classification of survey units is qualitative, while decision error rates, variability, and the lower bound of the gray region are quantitative measures.

The major activity in determining the usability of data based on survey activities is assessing the effectiveness of measurements. Scanning and direct measurements taken during survey activities and samples collected for analysis should meet site-specific objectives based on scoping and planning decisions.

Determining the usability of analytical results begins with the review of QC measurements and qualifiers to assess the measurement result and the performance of the analytical method. If an error in the data is discovered, it is more important to evaluate the effect of the error on the data than to determine the source of the error. The documentation described in Section N.2 is reviewed as a whole for some criteria. Data are reviewed at the measurement level for other criteria.

Factors affecting the accuracy of identification and the precision and bias of quantitation of individual radionuclides, such as calibration and recoveries, should be examined radionuclide by radionuclide. Table N.1 presents a summary of the QC measurements and the data use implications.

N.6.1 Precision

Precision is a measure of agreement among replicate measurements of the same property under prescribed similar conditions. This agreement is calculated as either the range or the standard deviation. It may also be expressed as a percentage of the mean of the measurements such as relative range (for duplicates) or coefficient of variation.

Table N.1 Use of Quality Control Data

Quality Control Criterion	Effect on Identification When Criterion is Not Met	Quantitative Bias	Use
Spikes (Higher than expected result)	Potential for incorrectly deciding a survey unit does not meet the release criterion (Type II decision error)	High	Use data as upper limit
Spikes (Lower than expected result)	Potential for incorrectly deciding a survey unit does meet the release criterion[a] (Type I decision error)	Low	Use data as lower limit
Replicates (Inconsistent)	None, unless analyte found in one duplicate and not the other—then either Type I or Type II decision error	High or Low[b]	Use data as estimate—poor precision
Blanks (Contaminated)	Potential for incorrectly deciding a survey unit does not meet the release criterion (Type II decision error)	High	Check for gross contamination or instrument malfunction
Calibration (Bias)	Potential for Type I or Type II decision errors	High or Low[b]	Use data as estimate unless problem is extreme

[a] Only likely if recovery is near zero.
[b] Effect on bias determined by examination of data for each radionuclide.

For scanning and direct measurements, precision may be specified for a single person performing the measurement or as a comparison between people performing the same measurement. For laboratory analyses, precision may be specified as either intralaboratory (within a laboratory) or interlaboratory (between laboratories). Precision estimates based on a single surveyor or laboratory represent the agreement expected when the same person or laboratory uses the same method to perform multiple measurements of the same location. Precision estimates based on two or more surveyors or laboratories refer to the agreement expected when different people or laboratories perform the same measurement using the same method.

The two basic activities performed in the assessment of precision are estimating the radionuclide concentration variability from the measurement locations and estimating the measurement error attributable to the data collection process. The level for each of these performance measures

should be specified during development of DQOs. If the statistical performance objectives are not met, additional measurements should be taken or one (or more) of the performance parameters changed.

Measurement error is estimated using the results of replicate measurements, as discussed in Chapter 6 for field measurements and Chapter 7 for laboratory measurements. When collocated measurements are performed (in the field or in the laboratory) an estimate of total precision is obtained. When collocated samples are not available for laboratory analysis, a sample subdivided in the field and preserved separately can be used to assess the variability of sample handling, preservation, and storage along with the variability in the analytical process, but variability in sample acquisition is not included. When only variability in the analytical process is desired, a sample can be subdivided in the laboratory prior to analysis.

Summary statistics such as sample mean and sample variance can provide as assessment of the precision of a measurement system or component thereof for a project. These statistics may be used to estimate precision at discrete concentration levels, average estimated precision over applicable concentration ranges, or provide the basis for a continual assessment of precision for future measurements. Methods for calculating and reporting precision are provided in *EPA Guidance for Quality Assurance Project Plans* (EPA 1997a).

Table N.2 presents the minimum considerations, impacts if the considerations are not met, and corrective actions for precision.

N.6.2 Bias

Bias is the systematic or persistent distortion of a measurement process that causes errors in one direction. Bias assessments for radioanalytical measurements should be made using personnel, equipment, and spiking materials or reference materials as independent as possible from those used in the calibration of the measurement system. When possible, bias assessments should be based on certified reference materials rather than matrix spikes or water spikes so that the effect of the matrix and the chemical composition of the contamination is incorporated into the assessment. While matrix spikes include matrix effects, the addition of a small amount of liquid spike does not always reflect the chemical composition of the contamination in the sample matrix. Water spikes do not account for either matrix effects or chemical composition of the contamination. When spikes are used to assess bias, a documented spiking protocol and consistency in following that protocol are important to obtaining meaningful data quality estimates.

Table N.2 Minimum Considerations for Precision, Impact if Not Met, and Corrective Actions

Minimum Considerations for Precision	Impact When Minimum Considerations Are Not Met	Corrective Action
Confidence level as specified in DQOs. Power as specified in DQOs. Minimum detectable relative differences specified in the survey design and modified after analysis of background measurements if necessary One set of field duplicates or more as specified in the survey design. Analytical duplicates and splits as specified in the survey design. Measurement error specified.	Errors in decisions to act or not to act based on analytical data. Unacceptable level of uncertainty. Increased variability of quantitative results. Potential for incorrectly deciding a survey unit does meet the release criterion for measurements near the detection limits (Type I decision error).	For Surveying and Sampling: Add survey or sample locations based on information from available data that are known to be representative. Adjust performance objectives. For Analysis: Analysis of new duplicate samples. Review laboratory protocols to ensure comparability. Use precision measurements to determine confidence limits for the effects on the data. The investigator can use the maximum measurement results to set an upper bound on the uncertainty if there is too much variability in the analyses.

Activity levels for bias assessment measurements should cover the range of expected contaminant concentrations, although the minimum activity is usually at least five times the MDC. For many final status surveys, the expected contaminant concentration is zero or background, so the highest activity will be associated with the bias assessment measurements. The minimum and maximum concentrations allowable in bias assessment samples should be agreed on during survey planning activities to prevent accidental contamination of the environment or an environmental level radioanalytical laboratory.

For scanning and direct measurements there are a limited number of options available for performing bias assessment measurements. Perhaps the best estimate of bias for scanning and direct measurements is to collect samples from locations where scans or direct measurements were performed, analyze the samples in a laboratory, and compare the results. Problems associated with this method include the time required to obtain the results and the difficulty in

obtaining samples that are representative of the field measurement to provide comparable results. A simple method of demonstrating that analytical bias is not a significant problem for scanning or direct measurements is to use the instrument performance checks to demonstrate the lack of analytical bias. A control chart can be used to determine the variability of a specific instrument and track the instrument performance throughout the course of the survey. Field background measurements can also be plotted on a control chart to estimate bias caused by contamination of the instrument.

There are several types of bias assessment samples available for laboratory analyses as discussed in Chapter 7. Field blanks can be evaluated to estimate the potential bias caused by contamination from sample collection, preparation, shipping, and storage.

Table N.3 presents the minimum considerations, impacts if the considerations are not met, and corrective actions for bias.

Table N.3 Minimum Considerations for Bias, Impact if Not Met, and Corrective Actions

Minimum Considerations for Bias	Impact When Minimum Considerations Are Not Met	Corrective Action
Matrix spikes to assess bias of non-detects and positive sample results if specified in the survey design. Analytical spikes as specified in the survey design. Use analytical methods (routine methods whenever possible) that specify expected or required recovery ranges using spikes or other QC measures. No radionuclides of potential concern detected in the blanks.	Potential for incorrectly deciding a survey unit does meet the release criterion (Type I decision error): if spike recovery is low, it is probable that the method or analysis is biased low for that radionuclide and values of all related samples may underestimate the actual concentration. Potential for incorrectly deciding a survey unit does not meet the release criterion (Type II decision error): if spike recovery exceeds 100%, interferences may be present, and it is probable that the method or analysis is biased high. Analytical results overestimate the true concentration of the spiked radionuclide.	Consider resampling at affected locations. If recoveries are extremely low or extremely high, the investigator should consult with a radiochemist or health physicist to identify a more appropriate method for reanalysis of the samples.

N.6.3 Accuracy

Accuracy is a measure of the closeness of an individual measurement or the average of a number of measurements to the true value (EPA 1997a). Accuracy includes a combination of random error (precision) and systematic error (bias) components that result from performing measurements. Systematic and random uncertainties (or errors) are discussed in more detail in Section 6.8.1.

Accuracy is determined by analyzing a reference material of known contaminant concentration or by reanalyzing material to which a known concentration of contaminant has been added. To be accurate, data must be both precise and unbiased. Using the analogy of archery, to be accurate one's arrows must land close together and, on average, at the spot where they are aimed. That is, the arrows must all land near the bull's eye (see Figure N.1).

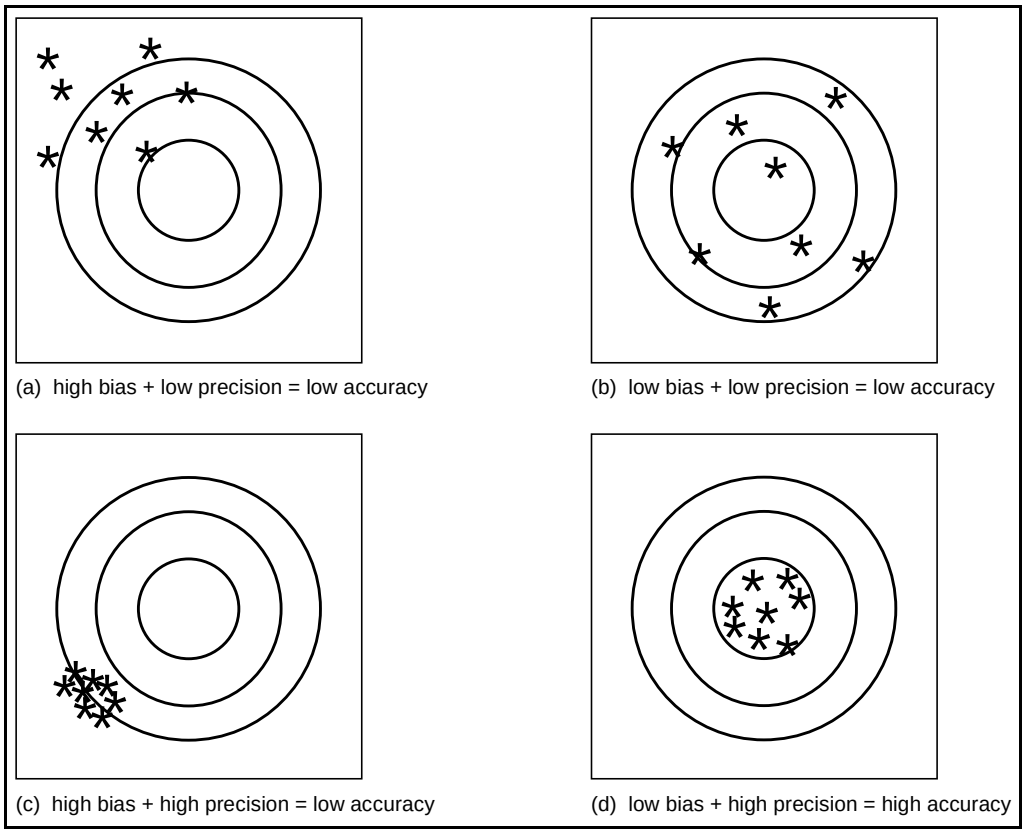

Figure N.1 Measurement Bias and Random Measurement Uncertainty

Accuracy is usually expressed either as a percent recovery or as a percent bias. Determination of accuracy always includes the effects of variability (precision); therefore, accuracy is used as a combination of bias and precision. The combination is known statistically as mean square error. Mean square error is the quantitative term for overall quality of individual measurements or estimators.

Mean square error is the sum of the variance plus the square of the bias. (The bias is squared to eliminate concern over whether the bias is positive or negative.) Frequently it is impossible to quantify all of the components of the mean square error—especially the biases—but it is important to attempt to quantify the magnitude of such potential biases, often by comparison with auxiliary data.

N.6.4 Representativeness

Representativeness is a measure of the degree to which data accurately and precisely represent a characteristic of a population parameter at a sampling point or for a process condition or environmental condition. Representativeness is a qualitative term that should be evaluated to determine whether *in situ* and other measurements are made and physical samples collected in such a manner that the resulting data appropriately reflect the media and contamination measured or studied.

Representativeness of data is critical to data usability assessments. The results of the environmental radiological survey will be biased to the degree that the data do not reflect the radionuclides and concentrations present at the site. Non-representative radionuclide identification may result in false negatives. Non-representative estimates of concentrations may be higher or lower than the true concentration. With few exceptions, non-representative measurements are only resolved by additional measurements.

Representativeness is primarily a planning concern. The solution to enhancing representativeness is in the design of the survey plan. Representativeness is determined by examining the survey plan. Analytical data quality affects representativeness since data of low quality may be rejected for use.

Table N.4 presents the minimum considerations, impacts if the considerations are not met, and corrective actions for representativeness.

N.6.5 Comparability

Comparability is the qualitative term that expresses the confidence that two data sets can contribute to a common analysis and interpolation. Comparability should be carefully evaluated to establish whether two data sets can be considered equivalent in regard to the measurement of a specific variable or groups of variables.

Table N.4 Minimum Considerations for Representativeness, Impact if Not Met, and Corrective Actions

Minimum Considerations for Representativeness	Impact When Minimum Considerations Are Not Met	Corrective Action
Survey data representative of survey unit. Documented sample preparation procedures. Filtering, compositing, and sample preservation may affect representativeness. Documented analytical data as specified in the survey design.	Bias high or low in estimate of extent and quantity of contaminated material. Potential for incorrectly deciding a survey unit does meet the release criterion (Type I decision error). Inaccurate identification or estimate of concentration of a radionuclide. Remaining data may no longer sufficiently represent the site if a large portion of the data are rejected, or if all data from measurements at a specific location are rejected.	Additional surveying or sampling. Examination of effects of sample preparation procedures. Reanalysis of samples, or resurveying or resampling of the affected site areas. If the resurveying, resampling, or reanalyses cannot be performed, document in the site environmental radiological survey report what areas of the site are not represented due to poor quality of analytical data.

Comparability is not compromised provided that the survey design is unbiased, and the survey design or analytical methods are not changed over time. Comparability is a very important qualitative data indicator for analytical assessment and is a critical parameter when considering the combination of data sets from different analyses for the same radionuclides. The assessment of data quality indicators determines if analytical results being reported are equivalent to data obtained from similar analyses. Only comparable data sets can be readily combined.

The use of routine methods (as defined in Section 7.6) simplifies the determination of comparability because all laboratories use the same standardized procedures and reporting parameters. In other cases, the decision maker may have to consult with a health physicist and/or radiochemist to evaluate whether different methods are sufficiently comparable to combine data sets.

There are a number of issues that can make two data sets comparable, and the presence of each of the following items enhances their comparability (EPA 1997a).

- two data sets should contain the same set of variables of interest.
- units in which these variables were measured should be convertible to a common metric.
- similar analytic procedures and quality assurance should be used to collect data for both data sets
- time of measurements of certain characteristics (variables) should be similar for both data sets
- measuring devices used for both data sets should have approximately similar detection levels
- rules for excluding certain types of observations from both samples should be similar
- samples within data sets should be selected in a similar manner
- sampling frames from which the samples were selected should be similar
- number of observations in both data sets should be of the same order of magnitude

These characteristics vary in importance depending on the final use of the data. The closer two data sets are with regard to these characteristics, the more appropriate it will be to compare them. Large differences between characteristics may be of only minor importance depending on the decision that is to be made from the data.

Table N.5 presents the minimum considerations, impacts if they are not met, and corrective actions for comparability.

N.6.6 Completeness

Completeness is a measure of the amount of valid data obtained from the measurement system, expressed as a percentage of the number of valid measurements that should have been collected (*i.e.*, measurements that were planned to be collected).

Completeness for measurements is calculated by the following formula:

$$\%Completeness = \frac{(Number\ of\ Valid\ Measurements)\ x\ 100}{Total\ Number\ of\ Measurements\ Planned}$$

Completeness is not intended to be a measure of representativeness; that is, it does not describe how closely the measured results reflect the actual concentration or distribution of the contaminant in the media being measured. A project could produce 100% data completeness (*i.e.*, all planned measurements were actually performed and found valid), but the results may not be representative of the actual contaminant concentration.

Table N.5 Minimum Considerations for Comparability, Impact if Not Met, and Corrective Actions

Minimum Considerations for Comparability	Impact When Minimum Considerations Are Not Met	Corrective Action
Unbiased survey design or documented reasons for selecting another survey design. The analytical methods used should have common analytical parameters. Same units of measure used in reporting. Similar detection limits. Equivalent sample preparation techniques. Analytical equipment with similar efficiencies or the efficiencies should be factored into the results.	Non-additivity of survey results. Reduced confidence, power, and ability to detect differences, given the number of measurements available. Increased overall error.	For Surveying and Sampling: Statistical analysis of effects of bias. For Analytical Data: Preferentially use those data that provide the most definitive identification and quantitation of the radionuclides of potential concern. For quantitation, examine the precision and accuracy data along with the reported detection limits. Reanalysis using comparable methods.

Alternatively, there could be only 70% data completeness (30% lost or found invalid), but, due to the nature of the survey design, the results could still be representative of the target population and yield valid estimates. The degree to which lack of completeness affects the outcome of the survey is a function of many variables ranging from deficiencies in the number of measurements to failure to analyze as many replications as deemed necessary by the QAPP and DQOs. The intensity of effect due to incompleteness of data is sometimes best expressed as a qualitative measure and not just as a quantitative percentage.

Completeness can have an effect on the DQO parameters. Lack of completeness may require reconsideration of the limits for decision error rates because insufficient completeness will decrease the power of the statistical tests described in Chapter 8.

For most final status surveys, the issue of completeness only arises when the survey unit demonstrates compliance with the release criterion and less than 100% of the measurements are determined to be acceptable. The question now becomes whether the number of measurements is sufficient to support the decision to release the survey unit. This question can be answered by constructing a power curve as described in Appendix I and evaluating the results. An alternative

method is to consider that the number of measurements estimated to demonstrate compliance in Chapter 5 was increased by 20% to account for lost or rejected data and uncertainty in the calculation of the number of measurements. This means a survey with 80% completeness may still have sufficient power to support a decision to release the survey unit.

Table N.6 presents the minimum considerations, impacts if the considerations are not met, and corrective actions for completeness.

Table N.6 Minimum Considerations for Completeness, Impact if Not Met, and Corrective Actions

Minimum Considerations for Completeness	Impact When Minimum Considerations Are Not Met	Corrective Action
Percentage of measurement completeness determined during planning to meet specified performance measures.	Higher potential for incorrectly deciding a survey unit does not meet the release criterion (Type II decision error). Reduction in power. A reduction in the number of measurements reduces site coverage and may affect representativeness. Reduced ability to differentiate site levels from background. Impact of incompleteness generally decreases as the number of measurements increases.	Resurveying, resampling, or reanalysis to fill data gaps. Additional analysis of samples already in laboratory. Determine whether the missing data are crucial to the survey.

N.6.7 Selection and Classification of Survey Units

Selection and classification of survey units is a qualitative measure of the assumptions used to develop the survey plan. The level of survey effort, measurement locations (*i.e.*, random vs. systematic and density of measurements), and the integrated survey design are based on the survey unit classification. The results of the survey should be reviewed to determine whether the classification used to plan the survey is supported by the results of the survey.

If a Class 3 survey unit is found to contain areas of contamination (even if the survey unit passes the statistical tests), the survey unit may be divided into several survey units with appropriate classifications, and additional surveys planned as necessary for these new survey units.

Class 3 areas may only require additional randomly located measurements to provide sufficient power to release the new survey units. Class 2 and Class 1 areas will usually require a new survey design based on systematic measurement locations, and Class 1 areas may require remediation before a new final status survey is performed.

If a Class 2 survey unit is determined to be a Class 1 survey unit following the final status survey and remediation is not required, it may not be necessary to plan a new survey. The scan MDC should be compared to the $DCGL_{EMC}$ to determine if the measurement spacing is adequate to meet the survey objectives. If the scan MDC is too high, a new scan survey using a more sensitive measurement technique may be available. Alternatively, a new survey may be planned using a new measurement spacing or a stratified survey design may be implemented to use as much of the existing data as possible.

N.6.8 Decision Error Rates

The decision error rates developed during survey planning are related to completeness. A low level of completeness will affect the power of the statistical test. It is recommended that a power curve be constructed as described in Appendix I, and the expected decision error rates compared to the actual decision error rates to determine if the survey objectives have been accomplished.

N.6.9 Variability in Contaminant Concentration

The variability in the contaminant concentration (both in the survey unit and the reference area) is a key parameter in survey planning, and is related to the precision of the measurements. Statistical simulations show that underestimating the value of σ (the standard deviation of the survey unit measurements) can greatly increase the probability that a survey unit will fail to demonstrate compliance with the release criterion.

If a survey unit fails to demonstrate compliance and the actual σ is greater than the σ used during survey planning, there are several options available to the project manager. If the major component of variability is measurement uncertainty, a new survey can be designed using a measurement technique with higher precision or a lower MDC to reduce variability. If samples were collected as part of the survey design, it may only be necessary to reanalyze the samples using a method with higher precision rather than collect additional samples. Alternatively, the number of measurements can be increased to reduce the variability.

If the variability is due to actual variations in the contaminant concentration, there are still options available. If there is a high variability in the reference area, it may be appropriate to demonstrate the survey unit is indistinguishable from background. NUREG 1505 (NRC 1997b) provides guidance on determining whether this test is appropriate and performing the statistical tests. If the variability is caused by different contaminant distributions in different parts of the site (*i.e.*, changing soil types influences contaminant concentrations), it may be appropriate to redefine the survey unit boundaries to provide a more homogeneous set of survey units.

N.6.10 Lower Bound of the Gray Region

The lower bound of the gray region (LBGR) is used to calculate the relative shift, which in turn is used to estimate the number of measurements required to demonstrate compliance. The LBGR is initially set arbitrarily to one half the $DCGL_W$. If this initial selection is used to design the survey, there is no technical basis for the selection of this value. This becomes important because the Type II decision error rate (β) is calculated at the LBGR.

For survey units that pass the statistical tests, the value selected for the LBGR is generally not a concern. If the survey unit fails to demonstrate compliance, it may be caused by improper selection of the LBGR. Because the number of measurements estimated during survey planning is based on the relative shift (which includes both σ and the LBGR), MARSSIM recommends that a power curve be constructed as described in Appendix I. If the survey unit failed to demonstrate compliance because of a lack of statistical power, an adjustment of the LBGR may be necessary when planning subsequent surveys.