6 EVALUATE THE SURVEY RESULTS

6.1 Introduction

The assessment phase of the data life cycle involves the interpretation of survey results. Interpretation of survey results is very straightforward when all of the data are below or all of the data are above the action level, and the correct decision regarding disposition of the M&E is obvious. In these cases very little data interpretation is required. However, formal statistical tests provide a valuable tool when the survey results are neither clearly above nor entirely below the action level. In either case, statistical tests always can be used to support the survey design in helping to ensure the quantity and quality of data meet the data quality objectives (DQOs) and measurement quality objectives (MQOs). Figure 6.1 illustrates the assessment phase of the data life cycle.

6.2 Conduct Data Quality Assessment

Data quality assessment (DQA) is a scientific and statistical evaluation that determines whether data are the right type, quality, and quantity to support their intended use (EPA 2006b). There are five steps in the DQA process:

1. Review the DQOs and survey design.
2. Conduct a preliminary data review.
3. Select the statistical test.
4. Verify the assumptions of the statistical test.
5. Draw conclusions from the data.

The effort applied to DQA should be consistent with the graded approach used to develop the survey design. More information on DQA can be found in *Data Quality Assessment: A User's Guide* (EPA QA/G-9R, EPA 2006b) and *Data Quality Assessment: Statistical Tools for Practitioner*s (EPA QA/G-9S, EPA 2006c). Data should be verified and validated as described in the quality assurance project plan (QAPP). Guidance on data verification and validation can be found in MARSSIM Section 9.3 and MARLAP Chapter 8. Guidance on developing a QAPP is available in EPA QA/G-5 (EPA 2002a) and MARLAP Chapter 4.

6.2.1 Review the Data Quality Objectives and Survey Design

The first step in the DQA process is a review of the DQO outputs used to develop the survey design to ensure they are still applicable. The review of the DQOs and survey design should also include the MQOs (e.g., measurement uncertainty, detectability, quantifiability). For example, if the data show the measurement uncertainty exceeds the estimate used to design the survey, the DQOs and MQOs should be revisited.

The survey design should be reviewed for consistency with the DQOs. For example, the review should verify that the appropriate number or amount of measurements were performed in the correct locations and were analyzed using measurement methods with adequate sensitivity.

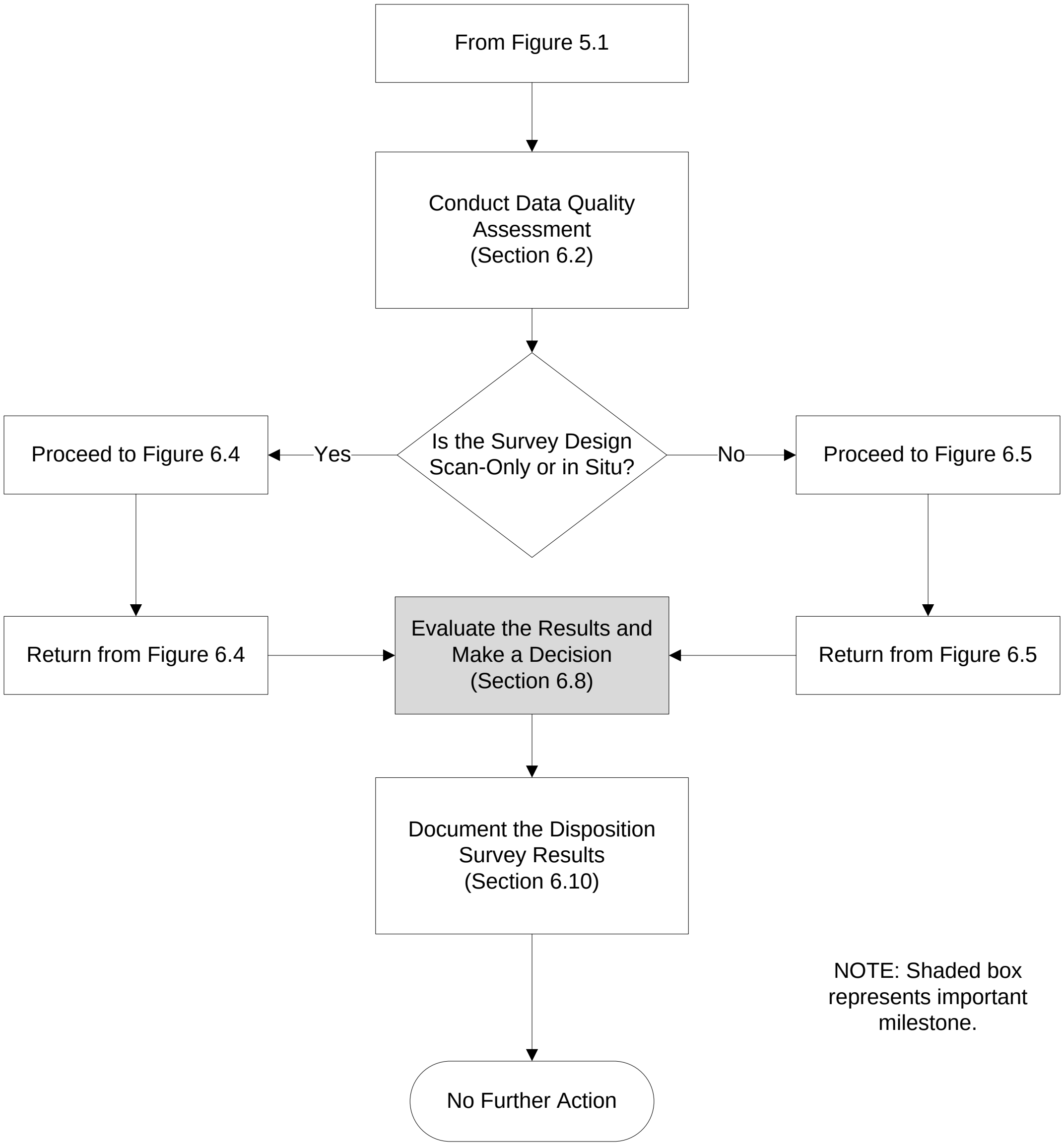

From Figure 5.1
Conduct Data Quality Assessment (Section 6.2)
Is the Survey Design Scan-Only or in Situ?
Yes
No
Proceed to Figure 6.4
Proceed to Figure 6.5
Return from Figure 6.4
Evaluate the Results and Make a Decision (Section 6.8)
Return from Figure 6.5
Document the Disposition Survey Results (Section 6.10)
NOTE: Shaded box represents important milestone.
No Further Action

Figure 6.1 The Assessment Phase of the Data Life Cycle

In cases where the survey did not involve taking discrete measurements or samples (i.e., scan-only, conveyor systems, or in situ surveys), it is imperative that the minimum detectable concentrations (MDCs) be calculated realistically and they truly reflect at least 95% probability that concentrations at or about the MDC were detected. Clearly, MDCs must be capable of detecting radionuclide concentrations or levels of radioactivity at or below the upper bound of the gray region (UBGR). When detection decisions are made for individual items (i.e., Scenario B) the MDC should be less than or equal to the UBGR.

The minimum quantifiable concentration (MQC) is defined as the radionuclide concentration or level of radioactivity at which the measurement method gives results with a specified relative standard deviation $1/k_Q$, where k_Q is usually chosen to be 10 (see Section 5.8, MARLAP Section

19.4.5, MARLAP Section 19.7.3). MARSAME recommends that the MQC should be no larger than the upper bound of the gray region (UBGR) when making quantitative comparisons of the mean survey data to the action level (i.e., Scenario A). This is an expression of the fact that the MQC, unlike the MDC used for a simple detection decision, addresses the relative uncertainty of the data value obtained. If the objective of the disposition survey is to quantify radionuclide concentrations near the UBGR, the MQC should be no larger than the UBGR.[1]

For MARSSIM-type surveys (Section 4.4.3) it is important to collect sufficient data to support a disposition decision. This is particularly important in cases where the radionuclide concentrations are near the action level. This can be done prospectively during survey design to test the efficacy of a proposed survey design (see Chapter 4), or retrospectively during interpretation of survey results to demonstrate the objectives of the survey design have been achieved. The procedure for generating power curves for the Sign test and the Wilcoxon Rank Sum test are provided in Appendix I of MARSSIM. Note that the accuracy of a prospective power curve depends on estimates of data variability and the planned number of measurements. After the data are analyzed, the sample standard deviation provides an estimate of data variability and the actual number of valid measurements are known, and these two parameters are used to generate a retrospective power curve (see MARSSIM Appendix I). The consequence of inadequate power is an increased Type II decision error rate. For Scenario A, this means M&E that actually meet the release criteria have a higher probability of being incorrectly determined not to meet the release criterion. For Scenario B, this means M&E that actually do not meet the release criterion have a higher probability of being incorrectly determined to meet the release criterion.

6.2.2 Conduct a Preliminary Data Review

A preliminary data review is performed to learn more about the structure of the data by identifying patterns, relationships, or potential anomalies. The preliminary data review includes reviewing quality assurance (QA) and quality control (QC) reports, performing a graphical data review, and calculating basic statistical quantities.

6.2.2.1 Review Quality Assurance and Quality Control Reports

Quality assurance reports describing data collection and reporting processes provide valuable information about potential problems with or anomalies in the data. EPA QA/G-9R (EPA 2006b) recommends a review of (1) data validation reports that document the data collection, handling, analysis, reduction, and reporting procedures; (2) QC reports from laboratories or field stations that document measurement system performance including data from blanks, replicates, spikes, standards, and certified reference materials, or other internal QC measures; and (3) technical systems reviews, performance evaluation audits, and audits of data quality including data from performance evaluation measurements. EPA QA/G-9R (EPA 2006b) also suggests paying particular attention to information that can be used to check assumptions made during survey design using the DQO process, especially any anomalies in recorded data, missing values, deviations from SOPs, or the use of nonstandard data collection methods (e.g., new, emerging, or “cutting edge” technology). Verification of instrument calibrations and review of MQOs are

[1] The UBGR is either the action level for Scenario A or the discrimination limit for Scenario B (see Section 4.2).

particularly important to disposition surveys. Periodic measurements must be made to ensure the measurement systems remain within acceptable calibration and control limits.

Quality control measurements are performed during implementation of the survey design to monitor performance of the measurement methods, identify problems, and initiate corrective actions when necessary. The evaluation of QC measurements used to control measurement methods is distinct from the evaluation from survey results. MARLAP Section 18.3 (“Evaluation of Performance Indicators”), Attachment 18A (“Control Charts”), and Attachment 18B (“Statistical Tests for QC Results”) provide information on the evaluation of quality control measurements.

Reviewing QA and QC reports is the only preliminary data review performed for surveys where individual measurements are not recorded (e.g., scan-only surveys with hand-held instruments). This increases the importance of the QA and QC reports and should be considered during survey planning to ensure data quality is adequate to meet the survey objectives.

6.2.2.2 Perform a Graphical Data Review

Preparing and evaluating graphs and other visual depictions of the data may identify trends in the data that go unnoticed using purely numerical methods. The graphical data review may include posting plots, frequency plots, quantile plots, or other methods for visually interpreting data. General guidance on performing a graphical data review and exploratory data analysis is provided in EPA QA/G-9R (EPA 2006b) and by the National Institute of Science and Technology (NIST 2006). A graphical data review cannot be performed unless the measurement results are recorded. Surveys where recording individual measurement results is not required (e.g., scan-only surveys with hand-held instruments) do not receive a graphical data review.

A posting plot is simply a map of the survey unit with the data values entered at the measurement locations. This type of plot potentially reveals heterogeneities in the data, especially possible clusters of elevated radionuclide concentrations. For a reference material survey a posting plot can reveal spatial trends in background data that might affect the results of the statistical tests. If the posting plot reveals systematic spatial trends in the M&E, the cause of the trends should be investigated. In some cases the trends could be attributable to residual radioactivity, but they may also be caused by inhomogeneities in the ambient background in the area the survey is performed. EPA QA/G-9S (EPA 2006c) provides additional diagnostic tools for examining spatial trends. The role of a posting plot for a conveyorized system would be a time series display of the data showing any trends between adjacent batches of M&E conveyed past the detector.

The geometric configuration of most M&E survey units composed of a few large irregularly shaped pieces of M&E is transitory. The arrangement of tools and piles of scrap metal, for example, changed as volumes of material were moved, or even as individual pieces were handled during the survey (Section 5.3). In these cases some identifying marks, numbers, or bar-code labels should have been used to identify and track where measurements were made, at least until it is determined that the M&E meet the disposition criteria. Such marking and labeling need not be permanent, but may be made with materials such as chalk or removable labels.

A frequency plot, or histogram, is a useful tool for examining the general shape of a distribution. This plot is a bar chart of the number of data points within a certain range of values. A frequency plot reveals any obvious departures from symmetry, such as skewness or bimodality (two peaks), in the data distributions for the M&E or reference material.

The presence of two peaks in the M&E data set frequency plot may indicate the presence of small areas of elevated activity. In some cases it may be possible to identify an appropriate background distribution within the M&E data set. This type of data interpretation generally depends on site-specific considerations and should only be pursued after consultation with the responsible regulatory agency.

The presence of two peaks in the M&E or reference material frequency plots may also indicate a mixture of materials with different intrinsic radiation backgrounds. The greater variability in the data caused by the presence of such a mixture reduces the power of the statistical tests. These situations should be avoided whenever possible through segregation of M&E (see Section 5.4) and carefully matching the reference materials to the M&E being surveyed.

When data are obtained from scan-only surveys incorporating data loggers, large quantities of data are usually recorded. In essence, 100% of Class 1 M&E are measured. While the survey coverage may be less than 100% for Class 2 and Class 3 M&E, the number of data points is still likely to be large. As long as there was no bias in the selection of areas that were scanned, the frequency plot will be close to the population distribution of radioactivity levels in the M&E. The mean and standard deviation calculated from these logged values should be very close to the corresponding population values.

For conveyorized survey monitors, the data may be interpreted batch-by-batch as it is scanned. In this case, the data treatment would be most similar to a single in situ measurement used to evaluate all of the M&E. If, on the other hand, the data were logged continuously the data treatment would be similar to a scan-only survey using data loggers.

6.2.2.3 Calculate Basic Statistical Quantities

Radiological survey data are usually obtained in units (e.g., counts per unit time) that have no intrinsic meaning relative to the action levels. For comparison of survey data to action levels, survey data from laboratory and field analyses are converted into action level units. MARSSIM Section 6.6 provides guidance on data conversion. Any uncertainty associated with data conversion should be included in the estimate of measurement uncertainty (Section 5.6). For surveys where individual results are not recorded (e.g., scan-only surveys with hand-held instruments) the uncertainty is associated with converting the action level into the units provided by the instrument in the field. Because individual results are not recorded, no statistical quantities can be calculated.

Basic statistical quantities that should be calculated for the sample data set include the mean, standard deviation, and the median. Other statistical quantities may be calculated based on the survey objectives.

Example 1: Suppose the following 10 measurement results are obtained from a disposition survey:

9.1, 10.7, 13.6, 3.4, 13.3, 7.9, 4.5, 7.7, 8.3, 10.4

The mean of the data (μ) is 8.89 and the standard deviation (σ) is 3.3231.
The next 10 measurement results are from an appropriate matching reference material:

6.2, 13.8, 15.2, 9.3, 6.7, 4.9, 7.1, 3.6, 8.8, 8.9

The mean of the reference data (μ) is 8.45 and the standard deviation (σ) is 3.6713.

The means of the two data sets can be compared to provide a preliminary indication of the survey unit status. [2] The difference is 0.44, with the M&E being investigated having a higher mean concentration. If the mean for the M&E exceeds the mean for the reference material by more than the action level, the M&E clearly do not meet the disposition criterion. On the other hand, if the difference between the largest M&E measurement (13.6 for this example) and the smallest reference material measurement (3.6 for this example) is below the action level, the M&E will pass the Wilcoxon Rank Sum (WRS) test (Section 6.6), but will have to meet other criteria as well.

The value of the sample standard deviation is especially important. If the standard deviation is too large compared to what was assumed for variability during development of the survey design, this may indicate an insufficient number of samples were collected to achieve the desired power for the statistical test. As previously mentioned, inadequate power can lead to an increase in the Type II decision error rate.

The median is the middle value of the data set when the number of data points is odd or the mean of the two middle values when the number of data points is even. A large difference between the mean and the median indicate a potential skew in the data. This would also be evident in a histogram of the data.

Examining other statistical quantities such as the maximum, minimum, and range may provide additional useful information. When there are 30 or fewer data points, range values greater than 4 or 5 standard deviations would be unusual.

Example 2: For the example M&E data set the minimum is 3.4 and the maximum is 13.6. The range is 13.6 − 3.4 = 10.2. The range is equal to 3.1 standard deviations (i.e., 10.2/3.3). Thus, the range for this example data set is not unusually large. The range may be greater for larger data sets.

[2] Note the use of significant digits in this example. Because all of the numbers in the text are interim values in calculating the difference between two means, they are not rounded. If the mean and standard deviation values were to be reported as results they would be rounded to two significant digits because the original data is a mixture of numbers with two and three significant digits. If the data were rounded after each calculation, the difference in the rounded means appears to be 0.4 (i.e., 8.9 minus 8.5), but the actual difference is 0.44 based on the un-rounded means (i.e., 8.89 minus 8.45). This is an example of how rounding numbers too early in the process can result in additional uncertainty.

6.2.3 Select the Statistical Tests

In most cases the selection of a statistical test is determined by the survey design used to collect the data. The most appropriate procedure for summarizing and analyzing the data is chosen based on the preliminary data review. If the preliminary data review indicates that the assumptions used to develop the survey design are valid, the statistical tests and evaluation methods determined should then be applied. If the assumptions used to develop the survey design are determined to be invalid, it may be necessary to consult a statistician to determine the most appropriate statistical test for evaluating the survey results.

6.2.3.1 Scan-Only Surveys

Scan-only surveys generate large amounts of data. Class 1 surveys measure all of the M&E. When less than 100 percent of the M&E are measured (i.e., Class 2 or Class 3 surveys) the areas that are measured are assumed representative of the areas that are not measured. This assumption should be checked during the preliminary data review (Section 6.2.2). The radionuclide concentrations or radioactivity in the areas that are not measured can be inferred based on the measurement results in the areas that are measured. Data indicating this inference may not be reasonable should result in re-evaluation of the survey design. For example, suppose the survey design specifies that ^{137}Cs is the radionuclide of concern and scanning 50% of the M&E is appropriate based on the expected distribution of radionuclide concentrations, expected levels of radioactivity, and the beta-gamma emissions from the radionuclide of concern. If additional historical data is found showing ^{239}Pu is also a radionuclide of concern, the survey design should be re-evaluated based on the presence of an alpha emitting radionuclide as well.

If disposition decisions will be made for individual items or based on individual measurement results, all of the results should be compared to the action level. Comparison to the action level based on a detection decision or measurement (Section 5.7) is discussed in Section 6.3. Individual measurement results can be recorded for scan-only surveys. The benefit of logging individual measurement results is the ability to statistically evaluate the data (e.g., calculate a mean and an upper confidence limit). If disposition decisions will be made based on the mean of logged data, an upper confidence limit for the mean is calculated and compared to the UBGR. This means that compliance with the disposition criterion can be demonstrated for the entire survey unit, even if some of the results exceed the UBGR. Evaluations using the upper confidence limit are discussed in Section 6.4. When less than 100% of the M&E are measured (i.e., Class 2 and Class 3 surveys), the total uncertainty includes both spatial and measurement uncertainty. Measuring 100% of the M&E (i.e., Class 1 survey) accounts for spatial variability, but there is still an uncertainty component resulting from variability in the measurement process.

Conveyorized systems that continually log the survey results also generate large amounts of data. An upper confidence limit for the mean can be used for the evaluation of data from these types of systems (see Section 6.4) in the same manner as logged scan data. Conveyorized systems that operate in a batch mode are essentially treated as single in situ measurements of small batches of M&E. The results generated by these types of systems are evaluated as a series of comparisons to the UBGR; using detection decisions based on the MDC (Section 6.3).

6.2.3.2 In Situ Surveys

In situ surveys may consist of a series of isolated measurements covering all or part of the M&E, a series of measurements with overlapping fields of view incorporating all (Class 1) or a portion (Class 2 or Class 3) of the M&E, or a single measurement incorporating all of the M&E (Section 4.4.2).

Similar to scan-only surveys, if disposition decisions will be made for individual items or based on individual measurement results, all of the results should be compared to the action level. Comparison to the action level based on a detection decision (Section 5.7) is discussed in Section 6.3. Unlike scan-only surveys, in situ surveys are likely based on a limited number of data points. To perform in situ measurements, assumptions were made about the distribution of radioactivity within the volume of M&E being measured. These assumptions are inherent in the calibration of in situ measurement systems and the validity of these assumptions determines the appropriateness of the measurement. It is important to account for uncertainty in these assumptions when calculating the MDC and to evaluate these assumptions using QC measurements performed during the survey. If there is uncertainty about the true MDC or critical value, use conservative values for the efficiency as described in Section 7.5.2.

6.2.3.3 MARSSIM-Type Survey Designs

MARSSIM-type survey designs generally are used when instrumentation for scan-only or in situ measurement surveys do not provide sufficient sensitivity (e.g., the MDC is greater than the UBGR). A statistically based number of measurements is used to provide an estimate of the mean activity in each survey unit, and scanning is used to identify small areas of elevated activity between sample locations.

The number of measurements is determined by the statistical test. In most cases the statistical tests used in MARSSIM are appropriate for Scenario A. The criteria for choosing between the Sign test and the WRS test are described in MARSSIM Section 8.2.3. In general, when the radionuclide is not present in background (or its background concentration is negligible compared to the action level) and radionuclide-specific measurements are made, the Sign test (Section 6.5) is used. Otherwise, the WRS (Section 6.6) test should be used. The Sign test is designed to detect whether there is radioactivity in the M&E above the action level. The WRS test is used to compare measurements of the M&E to measurements performed on the reference material.

When Scenario B is used, the statistical tests described in NUREG-1505 (NRC 1998a) generally are used. The Sign test and the WRS test are still used, but the application of the test is adjusted to account for the difference in the null hypothesis. When using Scenario B, there is a potential for the WRS test to miss non-uniform radioactivity (i.e., slightly elevated radionuclide concentrations or levels of radioactivity over a portion of the survey unit). Randomization of the M&E through mixing or homogenization can eliminate this possibility. If randomization is not practical, the Quantile test (Section 6.7) should be used to evaluate survey units when the WRS test fails to reject the null hypothesis.

The results of scanning measurements performed as part of a MARSSIM-type survey are evaluated using the elevated measurement comparison (EMC). The EMC is simply a comparison to an action level (see Section 6.3). The action level used for the EMC is the action level for small areas of elevated activity. If there is no action level for elevated activity, the scanning results are compared to the action level for the mean activity in the survey unit. Additional information on the EMC is available in MARSSIM Section 8.5.1 and NUREG-1505 Chapter 8 (NRC 1998a).

6.2.4 Verify the Assumptions of the Tests

An evaluation to determine the data are consistent with the underlying assumptions of the statistical tests helps to validate the use of a particular test. One may also determine that certain departures from these assumptions are acceptable when given the actual data and other information about the project. The nonparametric tests described in this chapter assume that the data from the M&E or the reference material consist of independent measurements from each distribution. The primary issue associated with the evaluation of scan-only and single in situ measurement survey data is the MDC or MQC as discussed in Section 6.2.1.

Asymmetry in the data can be identified using a histogram or a Quantile plot. Information on histograms and Quantile plots is provided in MARSSIM Appendix I and NUREG-1505 Section 4.2.2 (NRC 1998a). As discussed in Section 6.2.2.3, data transformations can sometimes be used to minimize the effects of asymmetry.

One of the primary advantages to using the nonparametric tests is that they involve fewer assumptions about the data than their parametric counterparts. If parametric tests are used (e.g., Student's t test) any additional assumptions made in using these tests should be verified (e.g., testing for normality). These issues are discussed in detail in EPA QA/G-9S (EPA 2006c).

One of the more important assumptions made in the survey design is that the number of measurements is sufficient to achieve the DQOs set for the Type I (α) and Type II (β) decision error rates. Verification of the power of the statistical tests ($1-\beta$) may be of particular interest. Methods for assessing power are discussed in Appendix I.9 of MARSSIM. If there is not reasonable assurance the DQOs have been achieved, additional investigations including repeating the survey may be needed. The planning team can develop survey designs cautiously to avoid unnecessary and potentially costly decision errors by—

- Estimating the potential data variability conservatively,
- Taking more measurements than suggested by the DQO process, and
- Estimating the MDCs conservatively.

In the absence of other data, each of these estimates could be multiplied by a safety factor of 1.2 (i.e., increase the estimate by 20%). Examples of assumptions and possible methods for evaluating and verifying these assumptions are summarized in Table 6.1.

Table 6.1 Issues and Assumptions Underlying the Evaluation Method

Evaluation Method	Issue	Verification Method	Survey Type
Compare single measurements to a limit (Section 6.3)	Verify the MDC and Measurement Uncertainty	Review the MDC Review QA/QC Reports Review IA and DQOs	Scan-Only In situ
Compare an upper confidence limit for the mean to a limit (Section 6.4)	Verify the MQC and Measurement Uncertainty	Review the Measurement Uncertainty Review QA/QC Reports Review IA and DQOs	Scan Only In situ
Statistical Tests (Sections 6.5, 6.6, 6.7)	Verify the Assumptions of the Statistical Test (e.g., spatial independence, symmetry, data variance, power)	Preliminary Data Review (e.g., posting plot, histogram, summary statistics, power curve)	MARSSIM-Type Survey

Verification of scan-only and in situ survey results focuses on the estimates of the MDC and MQC values used to design the survey. If the assumptions used to estimate these values are incorrect, the survey design may be invalid.

The first step in evaluating the MDC and MQC is to review the assumptions used to develop these values. In general, the key assumptions are made in determining the source and detector efficiencies. QA and QC reports should be reviewed to evaluate measurement performance (e.g., scan speed, source geometry, distance from M&E to the detector, non-uniform response of large area detectors). The description of the M&E from the IA should be compared to the assumptions used to develop the efficiency.

In some cases it may be possible to compare the survey results of multiple measurement techniques. For example, if there are multiple radiations associated with the M&E it may be possible to compare gamma measurement results to alpha or beta measurement results to verify the survey results. Direct measurements may provide more quantitative results for areas of elevated activity identified during scan-only surveys.

It may be possible to use an entirely different survey method to provide information to support verification of assumptions used to design a survey. For example, smears or surface scrapings can be used to verify the presence of radionuclides or radioactivity on the surface.[3]

In situ measurements or sample collection and analysis may be used to verify the results of scan-only survey designs. Care must be taken to ensure comparability of survey methods before evaluating the results to avoid generating conflicting results. For example, consider an in situ survey used to demonstrate the mean activity is less than the action level. A scan-only survey method is used to verify the results and identifies an area of elevated activity. This discrepancy in results warrants additional investigation of the small area of elevated activity. The additional investigation should determine if the activity in this area actually causes the mean activity to exceed the disposition criterion.

[3] This smear procedure does not rule out additional volumetric activity.

6.2.5 Draw Conclusions from the Data

The types of measurements performed on M&E are—

- Scans,
- In situ or direct measurements at discrete locations, and
- Samples collected at discrete locations.

Specific details for conducting the Sign test and the WRS test are provided in Sections 6.5 and 6.6, respectively. When the data clearly show that the M&E meets or exceeds the disposition criterion, the result is often obvious without performing the formal statistical analysis. This is the expected outcome for Class 2 and Class 3 surveys. Table 6.2 summarizes examples of circumstances leading to specific conclusions based on a simple examination of the data.

Table 6.2 Summary of Evaluation Methods and Statistical Tests

Evaluation Method or Statistical Test	Survey Result	Conclusion
Comparison to a Limit (AL=0) − Scenario B only − Results may or may not be recorded − Scan-only or In situ surveys	All measurements less than the critical value corresponding to the MDC (e.g., does not exceed alarm set point)	M&E meet the disposition criterion
	Any measurement exceeds the critical value corresponding to the MDC	M&E do not meet the disposition criterion
Comparison to a Limit (AL≠0) − Scenario A or B − Results not recorded − Scan-only or In situ surveys	All measurements less than the critical value corresponding to the UBGR	M&E meet the disposition criterion
	Any measurement exceeds the critical value corresponding to the UBGR	M&E do not meet the disposition criterion
Comparison to Upper Confidence Limit − Scenario A or B − Results must be recorded − Scan-only or In situ surveys	Upper confidence limit less than UBGR	M&E meet the disposition criterion
	Upper confidence limit greater than UBGR	M&E do not meet the disposition criterion
Sign Test − Radionuclide not in background − Nuclide-specific measurements − Scenario A or B − MARSSIM-type surveys	All measurements less than the action level	M&E meet the disposition criterion
	Mean greater then the action level	M&E do not meet the disposition criterion
	Any measurement greater than the action level and the mean less than the action level	Conduct Sign test (and elevated measurement comparison, if necessary)
Wilcoxon Rank Sum Test − Radionuclide in background − Nuclide non-specific measurements − Scenario A or B − MARSSIM-type surveys	Difference between maximum survey unit measurement and minimum reference area measurement is less than the UBGR	M&E meet the disposition criterion
	Difference of survey unit mean and reference area mean is greater than the action level	M&E do not meet the disposition criterion
	Difference between any survey unit measurement and any reference area measurement greater than the action level or the difference of survey unit mean and reference area mean is less than the action level	Conduct WRS test (and elevated measurement comparison, if necessary)

Table 6.3 Summary of Evaluation Methods and Statistical Tests (Continued)

Evaluation Method or Statistical Test	Survey Result	Conclusion
Quantile Test − Test for non-uniform radioactivity − Combine with WRS test − Scenario B only − MARSSIM-type surveys	Difference between maximum survey unit measurement and minimum reference area measurement is less than the UBGR	M&E meet the disposition criterion
	Difference of survey unit mean and reference area mean is greater than the action level	M&E do not meet the disposition criterion
	Difference between any survey unit measurement and any reference area measurement greater than the action level or the difference of survey unit mean and reference area mean is less than the action level	Conduct Quantile test (and elevated measurement comparison, if necessary)

6.3 Compare Results to the UBGR

When disposition decisions will be made about individual items, or decisions will be based on individual measurement results, each result (plus or minus a multiple of its combined standard uncertainty) will be compared to the action level (see MARLAP Appendix C.4). In practice, this means that any result that exceeds the critical value (S_C, see Section 5.7 and Section 7.5.1) when the minimum detectable level (S_D, see Section 5.7 and Section 7.5.2) equals the UBGR provides evidence that the result exceeds the UBGR.

For Scenario A, if all the results are less than the action level, then the mean and the maximum activity must also be below the action level. Thus, the radionuclide concentrations or levels of radioactivity associated with the M&E demonstrate compliance with the disposition criterion. For Scenario B when the action level is not zero or background, all of the results must be below the critical value corresponding to the MDC set equal to the UBGR. If the action level is zero or background, Scenario B must be used and any indication of the presence of radionuclide concentrations or radioactivity above background (i.e., above the discrimination level) would result in rejecting the null hypothesis. For this situation, any measurement result exceeding the critical value corresponding to the required MDC indicates the potential presence of radionuclides or radioactivity above background. This applies to single in situ measurements as well as series of in situ measurements.

If there is an action level based on small areas of elevated activity or the maximum allowable value, the individual results can be compared directly to the action level. This applies primarily to the evaluation of scanning results for MARSSIM-type surveys (i.e., the EMC), but may be applied to scan-only survey data as well.

6.4 Compare Results Using an Upper Confidence Limit

The use of the upper confidence limit (UCL) can apply to both Scenario A and B for scan-only or in situ surveys where individual results are recorded. When disposition decisions are made about the estimated mean of a sampled population, the assessment of the survey results is accomplished by comparing a UCL for the mean to the UBGR. For scan-only surveys where there are a large number of data points, a simple comparison of the mean activity to the UBGR may be sufficient.

If individual scan-only survey results are recorded, a non-parametric confidence interval can be used to evaluate the results of the disposition survey. Similarly, a confidence interval can be used to evaluate a series of in situ measurements with overlapping fields of view. A one-tailed version of Chebyshev's inequality or software (e.g., EPA's ProUCL software) can be used to evaluate the probability of exceeding the UBGR (i.e., using a UCL). The use of a UCL applies to both Scenario A (where the UBGR equals the action level) and Scenario B (where the UBGR equals the discrimination limit).[4]

6.4.1 Calculate the Upper Confidence Limit

Chebyshev's inequality calculates the probability that the absolute value of the difference of the true but unknown mean of the population and a random number from the data set is at least a specified value. That is, given a specified positive number (*n*), a mean (μ), and a random number from the data set (*r*), then the probability that $|\mu\text{-}r|$ is greater than or equal to *n* is equal to α. In addition, a one-tailed version of the inequality can be used to calculate a UCL for a data set that is independent of the data distribution (i.e., there is no requirement to verify the data are from a normal, lognormal, or any other specified kind of distribution) by letting the inequality equal the UCL, as described in the following steps:

1. Calculate the mean (μ) and standard deviation (σ) of the number of results (*n*) in the data set.
2. For Scenario A, retrieve the Type I error rate (α) used to design the survey.
3. Using Chebyshev's inequality, calculate the maximum UCL using equation 6-1:

$$\text{UCL} = \mu + \sqrt{\frac{\sigma^2}{n\alpha} - \frac{\sigma^2}{n}} \qquad (6\text{-}1)$$

4. For Scenario B, substitute the Type II error rate (β) used to design the survey for α in Equation 6-1.
5. If the maximum UCL is less than the UBGR, the survey demonstrates compliance with the disposition criterion (i.e., reject the null hypothesis for Scenario A or fail to reject the null hypothesis for Scenario B).

Chebyshev's inequality must be used with caution when there are very few points in the data set. This is because the population mean and standard deviation in the Chebyshev formula are being estimated by the sample mean and sample standard deviation. In a small data set from a highly skewed distribution, the sample mean and sample standard deviation may be underestimated if the high concentration but low probability portion of the distribution is not captured in the sample data set. EPA has issued guidance on calculating UCLs for exposure point concentrations (EPA 2002b).[5] Software for implementing EPA's guidance is available (EPA 2006d).

[4] In the case of Scenario B, if the action level is zero and the radionuclide of concern does not appear in background, any positive radionuclide-specific detection would result in a rejection of the null hypothesis that there is zero activity.

[5] In MARSAME, "exposure point concentration" is used to mean a conservative estimate of the mean radionuclide concentration(s) in or on M&E.

6.4.2 Upper Confidence Limit Example: Class 1 Concrete Rubble

This example illustrates the survey design for concrete rubble using 3 inch × 3 inch NaI(Tl) detectors mounted on a conveyorized survey system to measure ^{137}Cs. A pile of concrete rubble was loaded on the conveyor and passed beneath the detectors at a pre-determined speed. Each one-second count recorded by a detector corresponds to approximately 9,800 cm^3 of concrete rubble (i.e., a 5-cm thick disk with a 50-cm diameter). The following information was used to design the survey:

- The selected disposition option was clearance, using Scenario A with the null hypothesis that the residual radioactivity exceeds the action level.
- The IA indicated the concrete was potentially volumetrically contaminated prior to being converted to rubble.
- The concrete rubble had a maximum particle dimension of less than 0.5 cm.
- The average background count rate was estimated to be 38,000 cpm based on preliminary surveys of non-impacted concrete, and was used for the LBGR.
- The action level was set at 20,000 cpm above the average background count rate, so the UBGR was set at 58,000 cpm.
- The estimated standard deviation of background count rate is 2,500 cpm based on preliminary survey data.
- The Type I decision error rate was set at 0.10, or 10%.

The survey consisted of 9,616 independent, one-second measurements that were recorded using a data logger. The mean count rate for the survey was 39,252 cpm, with a standard deviation (σ) of 5,465 cpm. The standard deviation of the mean, σ_N.was calculated using the following equation:

$$\sigma_N = \frac{\sigma}{\sqrt{N}} = \frac{5{,}465\ cpm}{\sqrt{9{,}616}} = 55.7\ cpm \qquad (6\text{-}2)$$

As noted earlier, with such a large data set, one can expect that the sample mean and standard deviation should be fairly close to their population values. The minimum count rate was 30,080 cpm, and the maximum count rate was 72,805 cpm. Note that although the mean concentration is well below the action level, there are data points that exceed the action level. Thus, a test against an UCL for the mean is warranted. Figure 6.2 shows a frequency plot of the survey results.

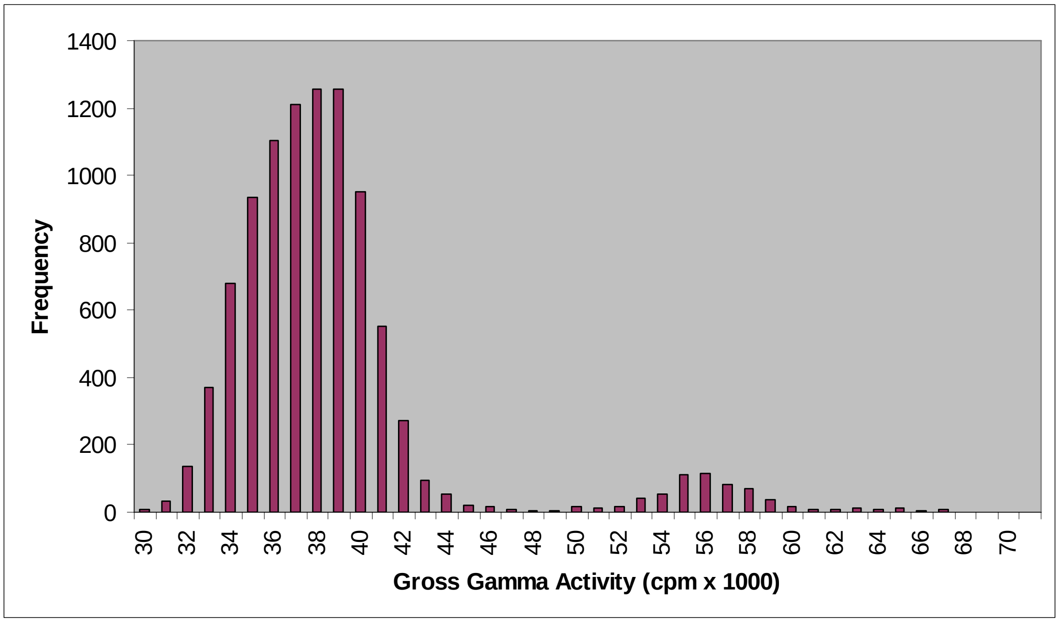


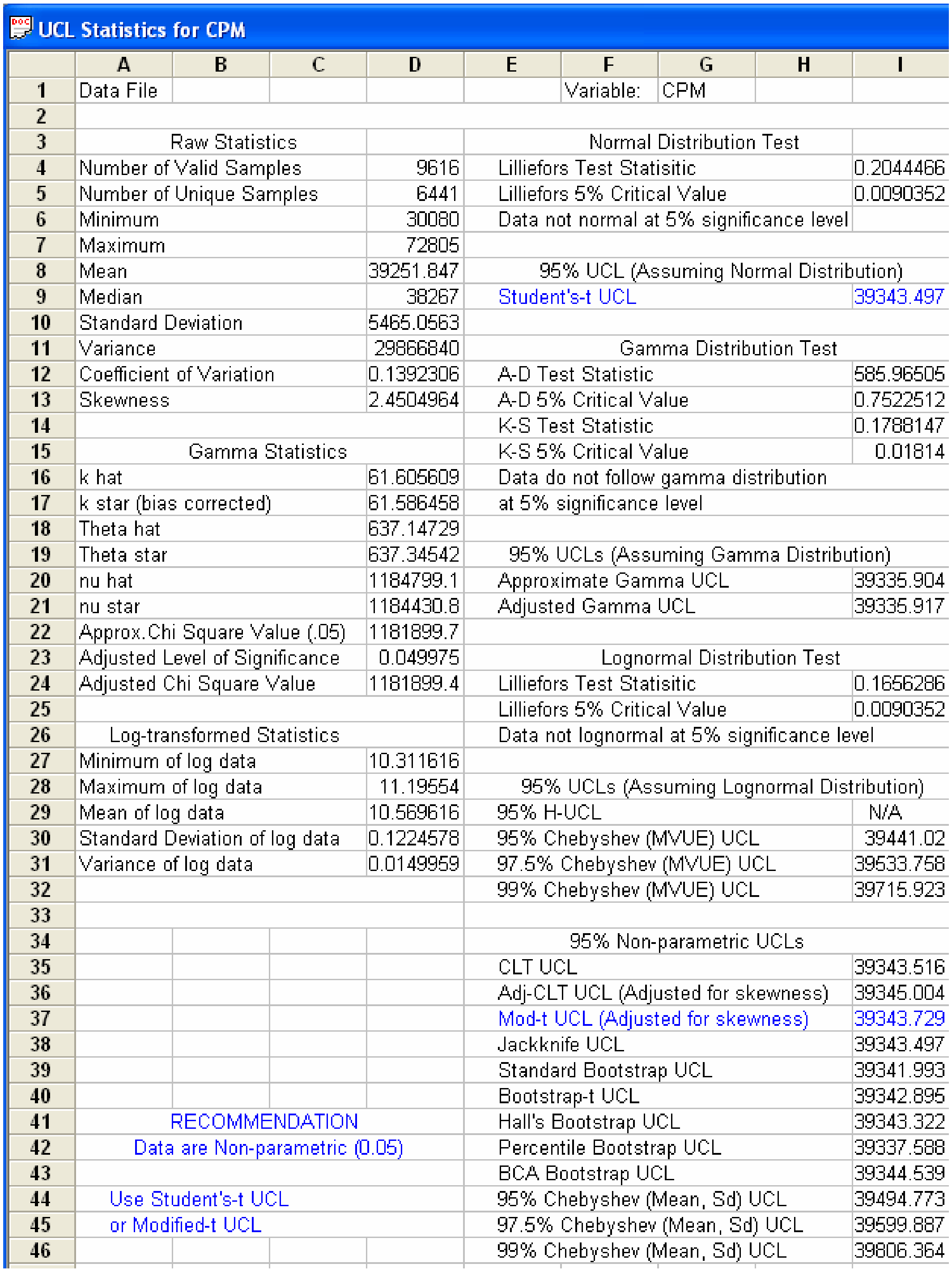
Figure 6.2 Frequency Plot of Concrete Rubble Data

If the sample size were small, however, the upper part of the bimodal distribution could be missed and the Chebyshev UCL could be underestimated. In this case, with a sample size of 9,616, the UCL was calculated using Equation 6-1 in Section 6.4.1.

$$\text{UCL} = 39{,}252 + \sqrt{\frac{(5{,}465)^2}{(0.10)(9{,}616)} + \frac{(5{,}465)^2}{(9{,}616)}} = 39{,}474\ \text{cpm} \qquad (6\text{-}3)$$

The UCL of 39,474 cpm is much less than the action level of 58,000 cpm. The null hypothesis that the level of radioactivity exceeds the disposition criterion is rejected.

The EPA ProUCL software was also applied to these data and the results are shown in Figure 6.3. The software has failed to find a good fit to the data for normal, lognormal or gamma distributions, which is hardly surprising given the bimodal nature of the data. The recommendation is that either a Student's *t* or a modified Student's *t* 95% UCL be used. These are both listed as about 39,343. These are lower than the 90% Chebyshev UCL of 39,474 used above, but that would not change the conclusion. A 95% Chebyshev UCL calculated according to Section 6.4.1 would have been 39,574. Note that the 95% Chebyshev UCL calculated by ProUCL, rounded to the nearest count, is slightly different, 39,495, because of the way that the sample mean and standard deviation are estimated before entering them in the Chebyshev formula. The ProUCL User's Manual can be consulted for details. However, with the number of data points at hand, there is little difference among any of the methods for computing an UCL.



UCL Statistics for CPM

	A	B	C	D	E	F	G	H	I
1	Data File					Variable:	CPM		
2									
3	Raw Statistics				Normal Distribution Test				
4	Number of Valid Samples			9616	Lilliefors Test Statisitic				0.2044466
5	Number of Unique Samples			6441	Lilliefors 5% Critical Value				0.0090352
6	Minimum			30080	Data not normal at 5% significance level				
7	Maximum			72805					
8	Mean			39251.847	95% UCL (Assuming Normal Distribution)				
9	Median			38267	Student's-t UCL				39343.497
10	Standard Deviation			5465.0563					
11	Variance			29866840	Gamma Distribution Test				
12	Coefficient of Variation			0.1392306	A-D Test Statistic				585.96505
13	Skewness			2.4504964	A-D 5% Critical Value				0.7522512
14					K-S Test Statistic				0.1788147
15	Gamma Statistics				K-S 5% Critical Value				0.01814
16	k hat			61.605609	Data do not follow gamma distribution				
17	k star (bias corrected)			61.586458	at 5% significance level				
18	Theta hat			637.14729					
19	Theta star			637.34542	95% UCLs (Assuming Gamma Distribution)				
20	nu hat			1184799.1	Approximate Gamma UCL				39335.904
21	nu star			1184430.8	Adjusted Gamma UCL				39335.917
22	Approx.Chi Square Value (.05)			1181899.7					
23	Adjusted Level of Significance			0.049975	Lognormal Distribution Test				
24	Adjusted Chi Square Value			1181899.4	Lilliefors Test Statisitic				0.1656286
25					Lilliefors 5% Critical Value				0.0090352
26	Log-transformed Statistics				Data not lognormal at 5% significance level				
27	Minimum of log data			10.311616					
28	Maximum of log data			11.19554	95% UCLs (Assuming Lognormal Distribution)				
29	Mean of log data			10.569616	95% H-UCL				N/A
30	Standard Deviation of log data			0.1224578	95% Chebyshev (MVUE) UCL				39441.02
31	Variance of log data			0.0149959	97.5% Chebyshev (MVUE) UCL				39533.758
32					99% Chebyshev (MVUE) UCL				39715.923
33									
34					95% Non-parametric UCLs				
35					CLT UCL				39343.516
36					Adj-CLT UCL (Adjusted for skewness)				39345.004
37					Mod-t UCL (Adjusted for skewness)				39343.729
38					Jackknife UCL				39343.497
39					Standard Bootstrap UCL				39341.993
40					Bootstrap-t UCL				39342.895
41	RECOMMENDATION				Hall's Bootstrap UCL				39343.322
42	Data are Non-parametric (0.05)				Percentile Bootstrap UCL				39337.588
43					BCA Bootstrap UCL				39344.539
44	Use Student's-t UCL				95% Chebyshev (Mean, Sd) UCL				39494.773
45	or Modified-t UCL				97.5% Chebyshev (Mean, Sd) UCL				39599.887
46					99% Chebyshev (Mean, Sd) UCL				39806.364

Figure 6.3 Screen Capture of Output from ProUCL Software for the Sample Data Set

6.5 Conduct the Sign Test

The Sign test is used to compare the measurement results from each survey unit with the applicable disposition criterion. The Sign test can be applied to either Scenario A or Scenario B. The Sign test should only be used if the radionuclide being measured is not present in background or if the radionuclide being measured is present at such a small fraction of the action level as to be considered insignificant. Otherwise, the WRS test described in Section 6.6 should be applied. Additional information on the Sign test can be found in Section 8.3 of MARSSIM and Chapter 5 of NUREG-1505 (NRC 1998a).

6.5.1 Apply the Sign Test to Scenario A

The Sign test is applied to Scenario A by counting the number of measurements from each survey unit that are less than the action level (i.e., UBGR). Each result is subtracted from the action level (AL – X_i), and the number of positive values is summed. The result is the test statistic $S+$. Discard any measurement that is exactly equal to the action level and reduce the sample size, N, by the number of such measurements. The value of $S+$ is compared to the critical values in A.3. If $S+$ is greater than the critical value (q) in the table, the null hypothesis is rejected.

6.5.2 Apply the Sign Test to Scenario B

The Sign test is applied to Scenario B in a manner similar to that used for Scenario A. However, for Scenario B the action level (i.e., LBGR) is subtracted from each result (X_i – AL), and the number of positive values is summed. The result is the test statistic $S+$. Discard any measurement that is exactly equal to the action level and reduce the sample size, N, by the number of such measurements. The value of $S+$ is compared to the critical values in Table A.3. If $S+$ is greater than the critical value (q) in the table, the null hypothesis is rejected.

6.5.3 Sign Test Example: Class 1 Copper Pipes

This example illustrates the disposition survey design for copper pipe sections using a gas-flow proportional counter to measure ^{239}Pu. Because the alpha background on the copper material is essentially zero, it was decided the Sign test would be used to determine whether the material meets the disposition criterion. The sample size was determined using the DQO Process and inputs such as the disposition option, action level, expected standard deviation of the measurement results, and the acceptable probability of making Type I and Type II decision errors.

The following inputs were used to develop the survey design–

- The selected disposition option was clearance.
- The survey was designed using Scenario A, with the null hypothesis that the residual radioactivity exceeds the action level.
- The IA indicated that the inside surfaces of the pipes potentially came in contact with liquids containing ^{239}Pu, but the outside surfaces were non-impacted.
- The gross activity action level was 100 dpm/100 cm^2. When converted to cpm the gross activity action level was 10 cpm (i.e., total efficiency = 0.10 counts per disintegration).
- The LBGR (i.e., the DL) was set at the expected activity level on the copper pipe sections (i.e., 5 net cpm, the same as the gross mean for an alpha background of 0).
- The standard deviation for the measurements was estimated at 2 cpm.

- The relative shift was calculated as (10−5)/2 = 2.5.
- The Type I and Type II decision error rates were both set at 0.05.

Table A.2a shows the number of measurements estimated to be needed for the Sign test, *N*, is 15 (α=0.05, β=0.05, and Δ/σ=2.5). Therefore, 15 surface activity measurements were randomly collected from the inside surfaces of the copper pipe sections. Survey results are shown in Table 6.3.

Table 6.3 Sign Test Example Data

Surface Concentration (cpm/100 cm^2)	**Surface Concentration (dpm/100 cm^2)**	**< Action Level?**
4	40	Yes
3	30	Yes
11	110	No
1	10	Yes
1	10	Yes
4	40	Yes
6	60	Yes
3	30	Yes
9	90	Yes
6	60	Yes
14	140	No
1	10	Yes
4	40	Yes
10	100	No
2	20	Yes
Number of measurements less than the action level (*S*+) = 12		

The surface activity values in Table 6.3 are determined by dividing the measured cpm by the total efficiency (0.10). No probe area correction is necessary. The mean count rate is 5 cpm, compared to the estimate of 5 cpm used for the LBGR, and the median is 4 cpm. The standard deviation is 4 cpm, which is higher than the value of 2 used to develop the survey design.[6] Thus, the power of the test is lower than planned. With the actual value of the relative shift (10−5)/4=1.2, 23 measurements should be collected.

With the 15 measurements collected, the actual Type II decision error rate is between 0.10 and 0.25 (the closest entries in Appendix A, Table A.2a are for α=0.05, β=0.10, and Δ/σ=1.2 with *N*=18, and α=0.05, β=0.25, and Δ/σ=1.2 with *N*=12). Three measurements exceed the action level. The portion of the material associated with these measurements merits further investigation using the elevated measurement comparison described in MARSSIM Section 8.5.1.

[6] Values are reported to one significant figure based on the data in Table 6.3. Interim calculations generally carry extra figures, so rounding to the appropriate number of significant figures only occurs for the final calculation. Rounding results too soon in the calculation may result in unnecessarily deleting individual results (i.e., when the result is exactly equal to the UBGR) resulting in lower statistical power.

The value of $S+$, 12, was compared to the appropriate critical value, q, in Appendix A, Table A.3. In this case, for N=15 and α=0.05, the critical value is 11. Because $S+$ exceeds q, reject the null hypothesis that the survey unit exceeds the action level. In this case, the slight loss of power attributable to underestimating the standard deviation did not affect the result. Pending the outcome of the investigation of the three elevated measurements, this survey unit has satisfied the disposition criteria established for clearance.

6.6 Conduct the Wilcoxon Rank Sum Test

The WRS test is used to compare each material survey unit with an appropriately chosen reference material. Each reference material should be selected on the basis of its similarity to the survey unit material, as discussed in Section 3.9. The WRS test can be applied to either Scenario A or Scenario B. Further information on the WRS test can be found in Section 8.4 of MARSSIM and Chapter 6 of NUREG- 1505 (NRC1998a).

6.6.1 Apply the WRS Test to Scenario A

The WRS test is applied to Scenario A as outlined in the following steps and further illustrated by the example in Section 6.6.2.

1. Obtain the adjusted reference material measurements, Z_i, by adding the action level to each reference material measurement, X_i. $Z_i = X_i +$ AL.
2. The m adjusted reference sample measurements, Z_i, from the reference material and the n sample measurements, Y_i, from the survey unit are pooled and ranked in order of increasing size from 1 to N, where $N = m + n$.
3. If several measurements are tied (i.e., have the same value), they are all assigned the mean rank of that group of tied measurements.
4. If there are t "less than" values, they are all given the mean of the ranks from 1 to t. Therefore, they are all assigned the rank $t(t+1)/(2t) = (t+1)/2$, which is the mean of the first t integers. If there is more than one MDC,[7] all observations below the largest MDC should be treated as "less than" values. If more than 40% of the data from either the reference material or the survey unit are reported as less than detectable, the WRS test *cannot* be used.
5. The sum of all the ranks, which is the sum of the first N positive integers, is $N(N+1)/2$, which equals W_r added to W_s. Thus, one needs only to sum the ranks of the either the adjusted reference measurements (W_r) or the sum of the ranks of the sample measurements (W_s).
6. Compare $\mathrm{W_r}$ with the critical value (q) given in Table A.4 for the appropriate values of n, m, and α. If W_r is greater than the tabulated value for q, reject the hypothesis that the survey unit exceeds the disposition criterion.

6.6.2 Apply the WRS Test to Scenario B

The WRS test is applied to Scenario B as outlined in the following steps:

[7] Examples of situations where there could be more than one MDC include using multiple laboratories to perform sample analyses and using different instruments with different backgrounds and different efficiencies to perform measurements.

1. Obtain the adjusted survey unit measurements, Z_i, by subtracting the LBGR from each survey unit measurement, Y_i. $Z_i = Y_i - \text{LBGR}$.
2. The n adjusted survey unit measurements, Z_i, and the m reference material measurements, X_i, are pooled and ranked in order of increasing size from 1 to N, where $N = m + n$.
3. If several measurements are tied (i.e., have the same value), they are all assigned the mean rank of that group of tied measurements.
4. If there are t "less than" values, they are all given the mean of the ranks from 1 to t. Therefore, they are all assigned the rank $t(t+1)/(2t) = (t+1)/2$, which is the mean of the first t integers. If there is more than one MDC, all observations below the largest MDC should be treated as "less than" values. If more than 40% of the data from either the reference material or the survey unit are reported as less than detectable, the WRS test *cannot* be used.
5. Sum the ranks of the adjusted measurements from the survey unit, W_s. The sum of all the ranks, which is the sum of the first N positive integers, is $N(N+1)/2$, which equals W_r added to W_s. Thus, one needs only to sum the ranks of the either the adjusted reference measurements (W_r) or the sum of the ranks of the sample measurements (W_s).
6. Compare $\mathrm{W_s}$ with the critical value (q) given in Table A.4 for the appropriate values of n, m, and α. (Note that when using this table for Scenario B, the roles of m and n are reversed. If the Quantile test is being used in addition to the WRS test, then $\alpha/\tilde{\d{2}}$ should be used rather than α.) If W_s is greater than the tabulated value for q, reject the hypothesis that the difference in the median concentration between the survey unit and the reference area is less than the LBGR.

6.6.3 WRS Test Scenario A Example: Class 2 Metal Ductwork

This example illustrates the use of the WRS test for releasing Class 2 metal ductwork. Assume that a gas-flow proportional detector was used to make gross (non-radionuclide-specific) surface activity measurements.

The DQOs from this survey unit include $\alpha = 0.05$ and $\beta = 0.05$, and the action level converted to units of gross cpm is 2,300 cpm, which is the UBGR. In this case, the WRS test is used because the estimated background level (2,100 cpm) was large compared to the action level. The estimated standard deviation of the measurements, σ, is 375 cpm. The estimated added activity level is 800 cpm; the LBGR is set at this value, and represents the DL. The relative shift is calculated as Δ/σ, which is (action level – LBGR)/σ, which equals 4.

The sample size needed for the WRS test can be found in Table A.2b for these DQOs. The result is nine measurements in each survey unit and nine in each reference material $\alpha = 0.05$, and $\beta = 0.05$, and $\Delta/\sigma = 4$). The ductwork was laid flat onto a prepared grid, and the 9 measurements needed in the survey unit were made using a random-start triangular grid pattern. For the reference materials, the measurement locations were chosen randomly on a suitable batch of material. Table 6.4 lists the gross count rate data obtained.

Table 6.4 Scenario A WRS Test Example Data

Data (cpm)	Area	Adjusted Data	Ranks	Reference Material Ranks
2180	R	4480	15	15
2398	R	4698	16	16
2779	R	5079	18	18
1427	R	3727	10	10
2738	R	5038	17	17
2024	R	4324	13	13
1561	R	3861	11	11
1991	R	4291	12	12
2073	R	4373	14	14
2039	S	2039	3	0
3061	S	3061	8	0
3243	S	3243	9	0
2456	S	2456	7	0
2115	S	2115	4	0
1874	S	1874	2	0
1703	S	1703	1	0
2388	S	2388	6	0
2159	S	2159	5	0
		Sum =	171	126

In the “Area” column, the code “R” denotes a reference material measurement and “S” denotes a survey unit measurement. The adjusted data were obtained by adding the action level to the reference material measurements (see Section 6.6.1, Step 1). The ranks of the data range from 1 to 18, because there are a total of 9+9 measurements (see Section 6.6.1, Step 2). Note that the sum of all of the ranks is still 18(18+1)/2 = 171. Checking this value with the formula in Step 5 of Section 6.6.1 is recommended to guard against errors in the rankings.

The total of the ranks belonging to the reference material measurements is 126. This is compared with the entry for the critical value of 104 in Table A.4 for $\alpha = 0.05$, with $n = 9$ and $m = 9$. Because the sum of the reference material ranks is greater than the critical value, the null hypothesis (i.e., that the mean survey unit concentration exceeds the action level) is rejected, and the ductwork is released.

This conclusion can be reached quickly by noting the difference between the largest survey unit measurement (3,243 cpm) and the smallest reference area measurement (1,427 cpm). This difference (3,243 – 1,427 = 1,816 cpm) is less than the action level of 2,300 cpm. Because the largest possible difference is less than the action level, the mean difference must also be less than the action level.

6.6.4 WRS Test Scenario B Example: Class 2 Metal Ductwork

This example illustrates the use of the Scenario B WRS test for releasing Class 2 metal ductwork, using the same data as in Section 6.6.3. The null hypothesis for Scenario B is that there is no detectable radioactivity above background.

In this case, the action level is set at no radioactivity detectable above the estimated background level (2,100 cpm). The LBGR is equal to the action level, and is set to zero. The regulator specified that the survey be able to detect an average excess of even 1,500 cpm being released. This value is the DL. The UBGR is set equal to the DL (i.e., 1,500 cpm), with $\beta = 0.025$. The owner of the ductwork felt that there was very little if any radioactivity above background present, and was willing to set $\alpha = 0.20$. The estimated standard deviation of the measurements, σ, was 375 cpm. The relative shift is Δ/σ = (UBGR – LBGR)/σ = (1,500 − 0)/375 = 4.

The sample size needed for the WRS test can be found in Table A.2b. The result is 9 measurements in each survey unit and 9 in each reference material $\alpha/2 = 0.10$, and $\beta = 0.025$, and $\Delta/\sigma = 4$. The data were obtained as in Section 6.6.3. Table 6.4 (on the previous page) lists the gross count rate data obtained. These data were reanalyzed using Scenario B and the results are shown in Table 6.5.

Table 6.5 Scenario B WRS Test Example Data

Data (cpm)	Area	Adjusted Data	Ranks	Survey Unit Ranks
2180	R	2180	11	0
2398	R	2398	13	0
2779	R	2779	16	0
1427	R	1427	1	0
2738	R	2738	15	0
2024	R	2024	6	0
1561	R	1561	2	0
1991	R	1991	5	0
2073	R	2073	8	0
2039	S	2039	7	7
3061	S	3061	17	17
3243	S	3243	18	18
2456	S	2456	14	14
2115	S	2115	9	9
1874	S	1874	4	4
1703	S	1703	3	3
2388	S	2388	12	12
2159	S	2159	10	10
		Sum =	171	94

In the “Area” column, the code “R” denotes a reference material measurement and “S” denotes a survey unit measurement. The adjusted data would be obtained by subtracting the LBGR from the survey unit measurements (see Section 6.6.2, Step 1), but because the LBGR is zero, no adjustment is needed. The ranks of the adjusted data range from 1 to 18, because there are a total of 9+9 measurements (see Section 6.6.2, Step 2). Note that the sum of all of the ranks is still 18(18+1)/2 = 171. Checking this value with the formula in Step 5 of Section 6.6.2 is recommended to guard against errors in the rankings. The total of the ranks belonging to the survey unit measurements is 94. This is compared with the entry for the critical value of 100 in Table A.4 for $\alpha = 0.10$, with $n = 9$ and $m = 9$. Because the sum of the reference material ranks is less than the critical value, the null hypothesis (i.e., that there is no detectable radioactivity above background) is not rejected, and the ductwork may be released if the Quantile test is passed.

6.7 Conduct the Quantile Test

The Quantile test was developed to detect differences between the surveyed M&E and the reference material that consist of a shift to higher values in only a fraction of the surveyed M&E. The Quantile test is only performed when Scenario B is used, and only if the null hypothesis is not rejected for the WRS test. Using the Quantile test, in tandem with the WRS test, results in higher power to identify M&E that do not meet the disposition criterion than either test by itself. Apply the Quantile test as follows:

1. Calculate α_Q ($\alpha_Q = \alpha/2$).
2. Obtain the adjusted survey unit measurements, Z_i, by subtracting the LBGR from each survey unit measurement, Y_i. $Z_i = Y_i$ - LBGR.
3. The n adjusted survey unit measurements, Z_i, and the m reference material measurements, X_i, are pooled and ranked in order of increasing size from 1 to N, where $N = m + n$.
4. If several measurements are tied (i.e., have the same value), they are all assigned the mean rank of that group of tied measurements.
5. Look up the values for r and q in Table A.5 based on the number of measurements in the survey unit (n), the number of measurements in the reference area (m), and α_Q. The operational decision described in the next step is made using the values for r and q.
6. If q or more of the r largest measurements in the combined ranked data set are from the survey unit, the null hypothesis is rejected.

This form of the Quantile test gives only approximate results, Because Table A.5 provides a limited number of combinations of n, m, and α_Q. It is recommended that several combinations of n, m, and α_Q be considered when interpreting the results of the Quantile test. Sections 7.2 and 7.3 of NUREG-1505 (NRC 1998a) provide additional guidance on interpreting the results of the Quantile test.

As an example, the Quantile test can be applied to the Class 2 Metal Ductwork example of section 6.6.4. Using $n = 9$, $m = 9$, and $\alpha_Q = 0.10$, the nearest entry in Table A.5d has for $r = 3$ $q = 3$ with $\alpha_Q = 0.105$ when $n = 10$ and $m = 10$. This means that all three of the highest measurement would have to be from the survey unit in order to reject the null hypothesis. From Table 6.5, one can see that the two largest measurements are from the survey unit, but the third largest is from the reference area. Because the ductwork has passed both the WRS and the Quantile test in the Scenario B example, one would conclude that it could be released from radiological controls.

6.8 Evaluate the Results: The Decision

Once the data and results of the tests are obtained, the specific steps required to make a disposition decision depends on the procedures approved by the regulator. The following considerations are suggested for the interpretation of the test results with respect to the disposition criteria. Note that the tests need not be performed in any particular order.

The interpretation of results from the data evaluation or statistical test is the decision to reject or not to reject the null hypothesis. For some of the survey designs the decision is straightforward, while for other designs the interpretation is more complex. Figures 6.4 and 6.4 summarize the interpretation of results.

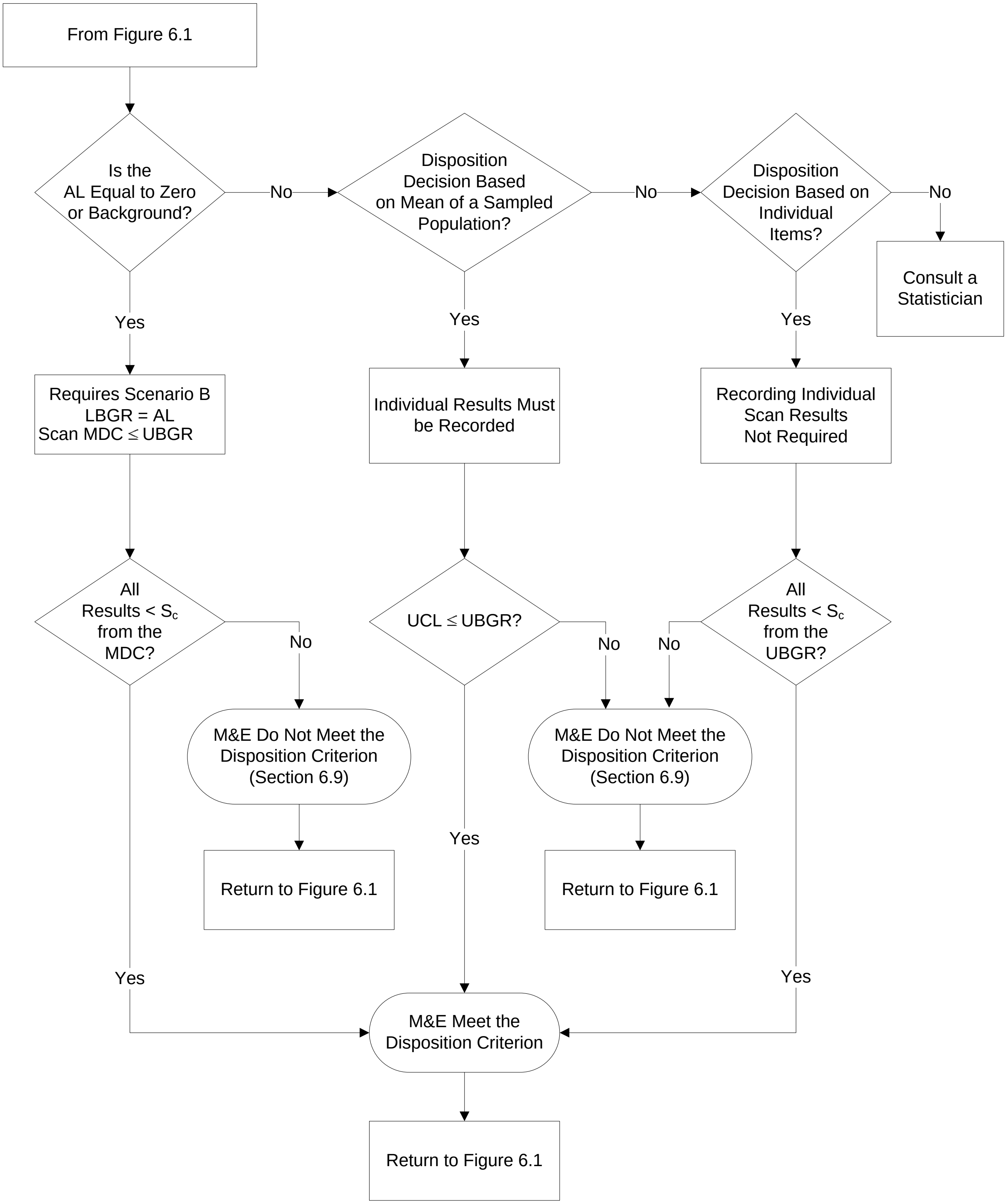
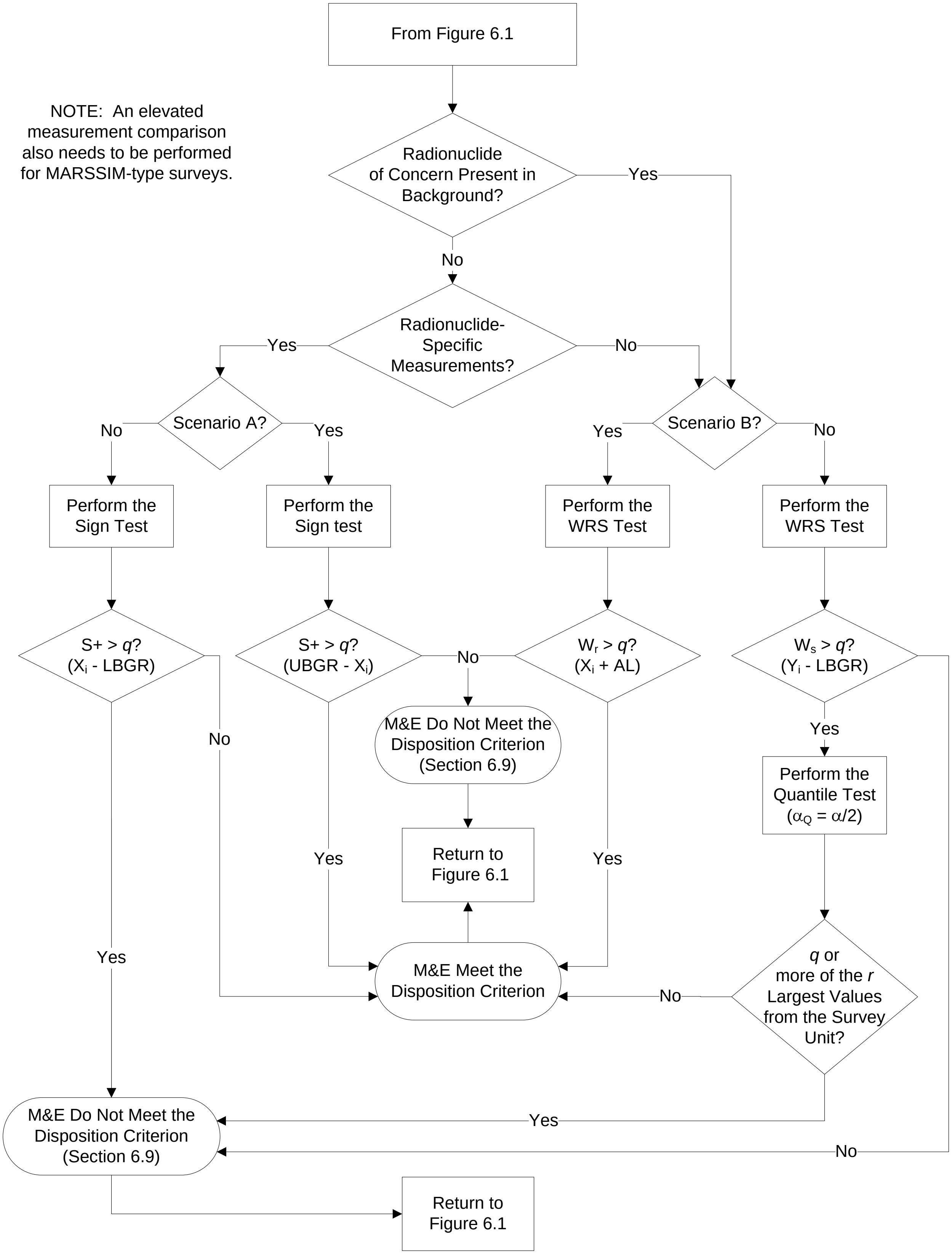

From Figure 6.1
Is the AL Equal to Zero or Background?
No
Disposition Decision Based on Mean of a Sampled Population?
No
Disposition Decision Based on Individual Items?
No
Consult a Statistician
Yes
Yes
Yes
Requires Scenario B LBGR = AL Scan MDC ≤ UBGR
Individual Results Must be Recorded
Recording Individual Scan Results Not Required
All Results < S_c from the MDC?
No
UCL ≤ UBGR?
No
No
All Results < S_c from the UBGR?
M&E Do Not Meet the Disposition Criterion (Section 6.9)
M&E Do Not Meet the Disposition Criterion (Section 6.9)
Return to Figure 6.1
Yes
Return to Figure 6.1
Yes
Yes
M&E Meet the Disposition Criterion
Return to Figure 6.1

Figure 6.4 Interpretation of Survey Results for Scan-Only and In Situ Surveys


From Figure 6.1
Radionuclide of Concern Present in Background?
Yes
No
Radionuclide-Specific Measurements?
Yes
No
Scenario A?
No
Yes
Scenario B?
Yes
No
Perform the Sign Test
Perform the Sign test
Perform the WRS Test
Perform the WRS Test
S+ > q? (X_i - LBGR)
S+ > q? (UBGR - X_i)
No
$W_r > q$? (X_i + AL)
$W_s > q$? (Y_i - LBGR)
No
Yes
M&E Do Not Meet the Disposition Criterion (Section 6.9)
Perform the Quantile Test ($\alpha_Q = \alpha/2$)
Return to Figure 6.1
Yes
Yes
Yes
M&E Meet the Disposition Criterion
No
q or more of the r Largest Values from the Survey Unit?
M&E Do Not Meet the Disposition Criterion (Section 6.9)
Yes
No
Return to Figure 6.1

NOTE: An elevated measurement comparison also needs to be performed for MARSSIM-type surveys.

Figure 6.5 Statistical Interpretation of Results for MARSSIM-Type Surveys

6.8.1 Compare Results to the UBGR

The process for interpreting results compared to the UBGR depends on the action level used to develop the survey design. Refer to Table 6.1 for issues and assumptions underlying this evaluation method.

If the action level is zero or background, Scenario B must be used:

- Compare every measurement result to the critical value corresponding to the required scan MDC.
- If all results are below the critical value, the M&E demonstrate compliance with the disposition criterion.
- Any results that exceed the critical value provide evidence of radionuclide concentrations or radioactivity levels exceeding the disposition criteria, so the M&E do not demonstrate compliance with the release criterion.

If the action level is not zero or background—

- Compare every measurement result to the critical value corresponding to the UBGR.
- If all results are below the critical value, the M&E demonstrate compliance with the disposition criterion.
- Any results that exceed the critical value provide evidence of radionuclide concentrations or radioactivity levels exceeding the disposition criteria, so the M&E do not demonstrate compliance with the release criterion.

Scan-only results are usually available as the data are collected. This real-time availability of results allows the surveyor to make decisions as the data are collected. M&E that exceed the action level can be identified and segregated during implementation of the survey. This “clean as you go” approach to surveys is only applicable for Class 1 surveys where there is high confidence in the quality and accuracy of detection decisions around the UBGR. Extensive documentation of the measurement process, previous applications of the process to the same or similar M&E, and verification of MDCs and MQCs is generally necessary to implement a “clean as you go” survey design.

6.8.2 Compare Results Using an Upper Confidence Limit

When decisions are made based on the mean of a sampled population, the survey results should be evaluated by comparison to a UCL (refer to Table 6.1 for issues and assumptions underlying this evaluation method):

- Compare every measurement result to the critical value corresponding to the UBGR.
- If all results are below the critical value, the M&E demonstrate compliance with the disposition criterion.
- If any results are above the critical value, calculate the UCL (Section 6.4.1).
- If the UCL is less than the UBGR, the M&E demonstrate compliance with the disposition criterion.

- If the UCL exceeds the UBGR, the M&E do not demonstrate compliance with the disposition criterion.
- Investigate measurements exceeding the UBGR.
- Results above the UBGR trigger a reevaluation of classification as Class 2.
- Results above the MDC trigger a reevaluation of classification as Class 3.

6.8.3 Compare Results for MARSSIM-Type Surveys

The process for evaluating MARSSIM-type survey results is more complicated. This process is explained in more detail in MARSSIM Section 8.5 (refer to Table 6.1 for issues and assumptions underlying this evaluation method):

- Calculate the test statistics (see Section 6.5.1, 6.6.1, 6.6.2, and 6.7).
- Look up the critical value in the appropriate statistical table in Appendix A.
- Evaluate the results of the statistical test as described in Figures 6.3 and 6.4.
- Evaluate individual results using the elevated measurement comparison (EMC).
- M&E must pass the statistical test and the EMC (if applicable) to demonstrate compliance.

If the null hypothesis is rejected under Scenario A, there is sufficient evidence to show the median radionuclide concentrations or radiation levels are below the disposition criterion. Under Scenario B, failing to reject the null hypothesis means there is insufficient evidence to overturn the initial assumption the M&E demonstrate compliance with the disposition criterion.

If the null hypothesis is rejected under Scenario B, additional investigations are required to determine the final disposition of the M&E (see Section 6.8.2). Failure to reject the null hypothesis under Scenario A also requires additional investigations.

6.9 Investigate Causes for Survey Unit Failures

When M&E fail to demonstrate compliance with the disposition criterion, the first step is to review and confirm the data that led to the decision. Once this is done, the DQO process can be used to evaluate potential problem areas leading to failure.

If the level of radioactivity on or in some Class 1 M&E exceeds the UBGR, the simplest solution might be to segregate those items for a different disposition decision. The concept of “clean as you go” for Class 1 M&E was discussed in Section 6.8.1 where individual objects or sample locations were identified during implementation of the survey design. A simple modification to this approach is to physically segregate the objects exceeding the action level as they are identified, or after reanalysis shows the cleaning was not effective. The segregated M&E can then be evaluated for a different disposition option (e.g., reuse, disposal).

Sometimes activity in excess of background can be removed from the M&E, or remediated, followed by re-evaluation or re-survey of the M&E. This approach may include evaluation of alternatives for remediation and a remedial action support survey prior to performing another final disposition survey.

If the radionuclides of concern have short half-lives, storage of the M&E until the radionuclides have decayed to acceptable levels, or "decay in place," may be an option. The planning team should consider the intrinsic value of the M&E along with storage and disposal costs when considering this option. When multiple radionuclides are present with significantly different half-lives (e.g., order of magnitude) radionuclide-specific measurements may be required to fully evaluate the acceptability of this option.

In other cases, a different disposition option (e.g., reuse, disposal) may be selected. If such a situation were encountered in evaluating Class 2 or Class 3 M&E, the classification would be questioned and the M&E would be reclassified and surveyed as Class 1 M&E. This may also bring other classification decisions into question.

As a general rule, it may be useful to anticipate possible modes of failure. These can be formulated as the problem to be solved using the DQO Process. Once the problem has been stated, the decision concerning the failing survey unit can be developed into a decision rule. For example, decide whether to attempt to remove the radioactivity or simply segregate certain types of M&E for low-level waste disposal. Next, determine the additional data, if any, needed to document that a survey unit where pieces with elevated measurements have been removed or areas of added activity removed demonstrates compliance with the disposition criterion. Alternatives to resolving the decision rule should be developed for each type of M&E that may fail the surveys. These alternatives can be evaluated against the DQOs, and a disposition survey design that meets the objectives of the project can be selected.

6.10 Document the Disposition Survey Results

Documentation of survey results is an important part of the disposition survey process. The form of this documentation can vary greatly depending on the survey objectives and regulatory or administrative requirements. Documentation of disposition survey results should be considered during survey design to ensure adequate records are provided during implementation. Generally, survey documentation requirements are provided as part of the documented survey design. Documented items may include—

- A description of the final disposition, such as disposal in a landfill, return to manufacture for refurbishment, sold as salvage, recycled as ferrous metal, etc.;
- A release statement to the transport carrier and recipient of the material indicating that the M&E described in the bill of laden meet(s) applicable state and federal regulations; and
- Results of QC measurements made during the conduct of release surveys and confirmation of compliance with facility SOPs and action levels.

In both routine and non-routine surveys, the documentation should comply with all applicable regulatory requirements. Development of survey documentation should allow for any necessary or required reviews.

If the disposition survey is a routine survey, then the survey will be documented as specified in the SOP. For example, routine surveys performed to clear M&E from a facility may require documentation that the instruments were calibrated and functioning properly and that trained personnel were on duty to perform the surveys. Quality assurance reviews and audits would be performed periodically (typically under a separate SOP) to document that the clearance surveys were being performed properly and that no M&E were cleared without first being surveyed. These records would document that properly trained personnel had adequately surveyed all M&E leaving the facility using properly functioning instruments. Documentation of individual measurement results may not be required or necessary.

If the survey is not routine, significantly more documentation may be required. This documentation should provide a complete and unambiguous record of the radiological status of the M&E relative to the selected action levels. In addition, sufficient data and information should be provided to enable an independent evaluation of the survey results, including repeating measurements at some future time Additional information on documentation is provided in Section 2.5, Section 3.6, Section 4.5, MARSSIM Sections 3.8 and 8.6, and MARSSIM Chapter 5.