CHAPTER 2
Preliminary Assessment and Site Investigation (PA/SI)

Section I
Preliminary Assessment

2.1. Introduction. A Preliminary Assessment (PA) is initiated after a CERCLA site (or suspected site) is identified. Statistical evaluations are not typically conducted for a PA. The purpose of the PA is to determine if a site poses a potential threat to human health or the environment. EPA maintains a list of actual and potential hazardous substance releases requiring CERCLA response. The property owner or agent is obliged to perform a PA; for Federal facilities, a PA is required within 18 months of listing (57 FR 31758; 17 July 1992).

2.1.1. The PA process collects information from existing resources. Generally, PA data are qualitative rather than quantitative, and do not require statistical evaluation. In some instances, historical chemical data may be available, but the PA does not require that such data be statistically manipulated. The EPA evaluates the site information according to the Hazard Ranking System (HRS) as detailed in 55 FR 51531 (14 December 1990). HRS calculations do not have statistical components. Some examples of PA information necessary to the HRS are as follows.

2.1.1.1. Identification of wastes or waste sources.

2.1.1.2. Physical site conditions, such as precipitation rates, depth to groundwater, or distance to surface water bodies.

2.1.1.3. Workers or residents at a site.

2.1.1.4. Local population within a set radius of a site.

2.1.2. Based on the results of the HRS, a site may warrant further investigation or no further action. Though quantitative statistical evaluations are not required during a PA, the following case study illustrates the value of a thorough qualitative evaluation of PA information.

2.2. Case Study 1—Examining Historical Data Sets. In the preliminary assessment of a landfill located on a manufacturing facility in Pennsylvania, some historical analytical data were available to the project team. The question raised, however, was whether or not those data would be usable in the PA. If the data were found to be usable and applicable, the landfill might be removed from further consideration in the CERCLA process. However, if the data were not found to be usable, then a Site Inspection (see Section II) would be needed. Moreover, if the data were used, prior to further validity testing (thus, explicitly assuming the data were reliable), and found later in the assessment to be erroneous, inaccurate and misleading conclusions would have been drawn.

2.2.1. Several different assessments of the data were required: i) Were the precision, accuracy, and representativeness of the data sufficient for the purpose? ii) Was the sampling design for the historical data sufficient for the purpose? and iii) Were the data comparable from historical event to historical event and could they be combined with new data, if necessary, to draw conclusions about the site?

2.2.2. The existing data were included in monitoring reports to the state. The reports consisted of little more than sample identification, date, and analytical results. Only positive detections were reported. Based on that information alone, the project team could not assess the quality of the data and concluded that unless additional information was obtained, the data could not be used as part of the PA. The site owners began to investigate the origins of the data.

2.2.3. In the interim, the project team assigned a geologist to examine the sampling design for the work. The facility had identified a single monitoring well, MW-02, as an upgradient location for comparison to a set of three downgradient wells, MW-03, MW-06, and MW-08. Through a review of well construction diagrams, as well as available topographic and hydrogeologic information, the geologist found that the well identified as upgradient was located within 3 feet of the landfill footprint, in a swale that received run-off from the landfill.

2.2.4. Thus, it was likely that the upgradient well was directly impacted by landfill operations and would not constitute an acceptable upgradient location. Further, MW-06 and MW-08 were found to be generally cross-gradient to MW-02 rather than directly downgradient, and that MW-03 had been screened in a perched aquifer, hydrologically isolated from the aquifer monitored by the other three wells.

2.2.5. Upon receipt of laboratory data packages for the historical data, the project team observed that a variety of different analytical methods and laboratories had been employed in the course of the work, resulting in mixed reporting limits and inconsistent detection of analytes. As a result of these assessments, the historical data were judged not to be usable for the PA.

2.2.6. In summary, prior monitoring appeared to indicate the presence of contamination (e.g., which would have triggered an RI), but additional evaluation data indicated that the data were not usable; therefore, an SI was initiated.

Section II
Site Inspection

2.3. Introduction. The Site Inspection (SI) is the next step in the CERCLA process. Statistical evaluations are often appropriate for an SI. Typically, the major objective of these evaluations is to establish the presence or absence of site contamination with respect to predefined decision limits. An SI is performed if the PA indicates the potential for hazardous materials to be present, if human or ecological receptors, or both, exist, and if there are potential complete exposure

pathways for the receptors. The SI generally focuses on establishing, through sampling and analysis, whether hazardous materials are present at concentrations that exceed some "screening criteria." The project planning team must establish decision limits or screening criteria prior to sampling and analyses. Generally, decision limits fall into the following categories:

2.3.1. Naturally occurring or known background levels (site-specific background information is typically unavailable at the SI stage).

2.3.2. Ecological benchmarks, which are dependent on analytes and media (typically developed with regulatory input).

2.3.3. Risk-based screening criteria for human health such as EPA Region IX Preliminary Remediation Goals (PRGs) or EPA Region III Risk-based Concentrations (RBCs) are available at the following Web sites.

http://www.epa.gov/region09/waste/sfund/prg/index.html

http://www.epa.gov/reg3hwmd/risk/index.htm

2.3.4. Applicable or relevant and appropriate requirements (ARARs). For example, Maximum Contaminant Levels (MCLs) for drinking water may be ARARs for some CERCLA sites.

2.3.5. During the DQO process, stakeholders identify the study questions, such as the presence or absence of contamination with respect to a set of decision limits, the nature and quantity of the data required to support the decision-making process, and the acceptable tolerances for decision errors. Selecting the screening criteria is *critical* for establishing both data quality objectives (DQO) and measurement quality objectives (MQOs). MQOs are established after DQO development. MQOs for analytical sensitivity must be adequate to report *quantitative* contaminant concentrations at levels less than the project decision limits. (Refer to Appendix G for a discussion of detection limits and quantitation limits.)

2.3.6. Team members must establish the DQOs for the project at the outset of the SI. In an SI, stakeholders must identify the problem at the site and how it will be evaluated, identify the decisions to be made using the data, and specify limits on that decision error. These will lead the project team to an optimal sampling design at a site. Appendix G discusses detection limits, quantitation limits, and censored data. Understanding the concepts in the context of ARARs guides part of the project planning.

2.4. Sampling Design. In general, statistical sampling designs are required to support statistical evaluations. Professional judgment, site-specific information, and DQOs must be used to select

the type of the statistical sampling design (e.g., *random*[*] as opposed to systematic sampling) and the required number of samples. The sampling design depends on factors such as the nature and distribution of the contamination in the study area, sampling cost, tolerances for decision error, and perceived level of decision uncertainty. For example, a small number of samples during the SI stage may be beneficial for short term cost considerations, but may not be adequate to achieve the desired tolerances for decision uncertainty and error and may, therefore, not be a cost-effective strategy by project closeout (as multiple sampling events rather than a single sampling event would typically be required to support decision-making).

2.4.1. Decision uncertainty refers to statistical variability, subjective judgment, randomness in the process, disagreement, and even imprecise wording inherent in the decision-making process (Moser 2000). Decision uncertainty is a function of the variability of the contaminant of concern in a study area and depends on the number of samples collected. For example, if the sample mean, $\bar{x}$, is an appropriate measure of site-wide contamination and the standard deviation of the sample mean, $s_{\bar{x}}$, measures the variability around $\bar{x}$, then the variability (and uncertainty) decreases as the number of samples n increases, because $s_{\bar{x}} = s/\sqrt{n}$. (Increasing the physical size of each sample would also decrease the variability.) It should also be noted that, in addition to decreasing the variability, $\bar{x}$ becomes a more accurate estimate of the population mean, μ, as n increases.

2.4.2. Site-specific information must be taken into account when selecting the sampling design. In particular, the team members need to identify potential source areas and any stratification they may represent. For example, suppose there are two sources of lead at a bomb reconditioning facility—stack emissions affecting surface soil and old buried waste piles affecting subsurface soil. This information can be used to design a sampling scheme for the "surface soil stratum" and a separate scheme for the "subsurface soil stratum." Likewise, there may be different study objectives for each stratum. Surface lead may be of concern for exposure of site workers and subsurface lead may be of concern for protection of groundwater. Stakeholders would need to identify these issues during project planning to develop an optimal site-wide sampling design.

[*] Appendices C and D.

2.4.3. Several different types of sampling designs are listed below. Appendix C presents a detailed explanation of these designs.

- Judgmental sampling.
- Random sampling.
  - Simple random sampling.
  - Stratified random sampling.
  - Systematic and grid sampling.
- Ranked set sampling.
- Adaptive cluster sampling.
- Composite sampling.

2.4.4. The TPP and DQO processes are used to develop an appropriate sampling design for the SI phase. Two case studies are presented below to illustrate sampling designs commonly used for SI.

2.5. Case Study 2—Judgmental Sampling, Oil/Water Separator. Project planners found an oil/water separator buried underground at a pipe mill. There was evidence of leakage to the surface soils around the tank and a release to groundwater was suspected. The objective was to determine if there was a measurable presence of oil floating on the water table.

2.5.1. Historical information and local knowledge allowed a hydrogeologist to determine the direction of groundwater flow. The hydrogeologist also knew of two monitoring wells in the area. One well was located upgradient to the separator; the second was cross-gradient.

2.5.2. The project planners decided to place a new monitoring well downgradient of the separator. Because they were looking for an oil product, the soil boring for the monitoring well was logged by a geologist who could then identify the water table depth. The well was installed so that the screen intersected the water table, where floating oil would most likely be visually detected.

2.5.3. Judgmental sampling was predominantly used in this example because the planners possessed significant existing site information. They knew the physical properties of the oil, they knew the hydrogeology of the site, and they were answering a nonquantitative question.

2.5.1. Case Study 4 predominantly illustrates the application of *composite sampling*[*] and *stratification*[†] for a SI, and the iterative nature of the DQO process when optimizing a sampling design.

[*] Appendices C and D.
[†] Appendix D.

2.6. Case Study 3—Arsenic Contamination in Soil. At an active manufacturing site, arsenic contamination was widespread in surface soils. Preliminary screening analyses and risk assessments identified worker exposure as the most likely concern. The site was initially divided (stratified) into 90 subunits related to work areas for a more in-depth evaluation of risk. Based on financial constraints, the project team was allocated a budget of $50,000 for SI sampling and analytical testing.

2.6.1. The aggregate initial cost of a field grab sample was $175, with $100 attributed to field collection and $75 attributed to laboratory analysis. The expected percent relative standard deviation (%RSD) for the analytical (laboratory) measurements was 5%. The estimated standard deviation, *s*, for the analytical method, at the decision limit of 600 ppm, was computed as 5% of 600 ppm or 30 ppm.

2.6.2. The planning team estimated the field component of the variability to be 10 times greater than the laboratory component of the variability. Thus, the %RSD for the field component of the variability was calculated by multiplying the %RSD for the analytical measurements by 10 (yielding a field component %RSD of 50%). This estimate was then multiplied by 600 ppm to yield a value of *s* equal to 300 ppm for the field component of variability (i.e., 50% of 600 ppm). The estimates for field and analytical variability (i.e., variance or s^2) were then combined and the standard deviation was calculated (s = 330 ppm). The maximum observed arsenic concentration was 720 ppm. The analytical method was deemed appropriate by the planning team. If historical sampling data were available, the data would be used to estimate the field variance and to test for normality.

2.6.3. The planning team principally considered two sampling design alternatives—simple random sampling and composite sampling (see Appendix C for a review of each sampling method). A *t*-test was used to calculate the sample size for simple random sampling (Appendix F). Given a decision error limit of $\alpha = 0.01$, more than 200 samples per work area would have been required (refer to Appendix L for a review of methods involved in setting and testing hypotheses). The total cost of this sampling effort would have exceeded $3 million.

2.6.4. Using similar methods, the team explored composite sampling, which would have required 30 samples to be collected per work area for a cost of over $1 million. Given the considerable cost burdens for both proposed sampling designs, the team decided to return to Step 6 of the DQO process and modify the decision error limits. The team found that by increasing α to 0.05, the composite sampling design would require the collection of 13 samples for each of the 90 work areas. This revised design had a total cost of $204,750, approximately one-fifth of the original estimate.

2.6.5. The team realized that they would have to find other means of generating an appropriate design while remaining within budget. To do this, the project team redefined the boundaries of the study (by revisiting Step 4 of the DQO process). The team recognized that one of the

drivers of the cost was the large number of separate study units (previously, the calculated sample size was applied to each of the study units). The planning team used exposure information for the contaminant to map out the potential or expected pathways in the surface soils through which the contaminant could spread. The potential pathways were categorized into four distinct spatial units.

2.6.6. Rather than collect data and make decisions for each of the 90 individual work areas, the team decided to sample and make decisions for each of the four risk areas. Recognizing that these larger areas carried greater decision error consequences, the team revisited Step 6 of the DQO process and established new limits for decision errors applicable to the four risk areas. The team established different decision confidence limits for each and recalculated the number of samples required. The cost of implementing this design was $38,850, which fell within the $50,000 budget for the sampling and analysis.

2.7. General Review of Sample Size Determination[*]. For typical statistical sampling designs, there are well-defined relationships between the number of required samples (i.e., sample size), tolerance for decision errors, and inherent variability of the analytical measurements and the target environmental population. One such relationship states that the sample size increases as the tolerance for decision error decreases or the variability increases. The sample size must be equal to or greater than the sample size required to achieve predetermined tolerances for decision errors. When confidence limits for the mean are of interest, an appropriate sample size is required to generate a sufficiently precise estimate of the true mean concentration of a chemical contaminant (refer to Paragraph 3.11 and Appendix K for additional discussion of confidence limits). For the example presented above, the sample size must be adequate to demonstrate that the upper limit of the CI for μ is less than the applicable regulatory threshold, RT. The required sample size must increase as s^2 increases and as the difference Δ (RT $-$ $\bar{x}$) decreases. In a well-conceived sampling plan for a solid waste, every effort should be made to estimate the values of $\bar{x}$ and s^2 before sampling starts. Case Study 3 illustrated that decision confidence affects sample size. Case Study 4 illustrates this concept in a different setting.

2.8. Case Study 4—Effect of Decision Confidence on Sample Number. Upon promulgation of the Toxicity Characteristic Leaching Procedure (TCLP) rule, a steel mill in Maryland contracted with a consultant to collect samples from various waste streams within the facility for TCLP analysis of metals (this case study considers only the cadmium data). One such waste stream was from a wastewater treatment system and consisted of collected sludges. Although no previous analysis of sludges had been done, cadmium had been monitored in the wastewater stream before treatment. The project manager believed that the wastewater data would be sufficient for establishing routine variability of cadmium in the sludge, assuming there were no great differences in the treatment process over time and a 10 times concentration factor from wastewater to sludge.

[*] Appendix C.

2.8.1. The project manager decided to use the past year's wastewater data to make preliminary estimates of the number of samples required to meet the statistical confidence requirements of the TCLP rule (i.e., $\alpha = 0.2$). Four results (in milligrams per liter [mg/L]) were available from the previous year as follows: 14.2, 9.6, 21.7, and 19.3.

2.8.2. The mean and variance of the results (as adjusted for concentration to sludge) were the following: $\bar{x} = 1.6$ mg/L and $s^2 = 2.2$ mg/L, respectively. The proposed water regulatory threshold value (RT) was 1 mg/L. Using the formula for simple random sampling, the project manager calculated the number of samples required as follows:

$$n = (t^2 \times s^2) \div (\mathrm{RT} - \bar{x})^2$$

where: n = number of samples required
t = Student's value for n–1 degrees of freedom and 0.8 confidence
s^2 = sample variance
$\bar{x}$ = sample mean
RT = regulatory threshold.

2.8.3. Thus, $n = [(0.9785)^2 \times 2.2]/(1 - 1.6)^2 = 6$ samples. Samples are an integer value, and should be reported without decimal fractions. (The value of t may be obtained from Table B-23, where $df = 3$ and $p = 0.8$.) Assuming a sampling cost of $50 per sample and an analytical cost of $25 per sample, this testing would cost $450.

2.8.4. The client's attorneys asked what the effect would be should they wish to establish a safety margin by increasing the decision confidence to $\alpha = 0.05$. The revised plan would require

$n = [(2.353)^2 \times 2.2]/(1 - 1.6)^2 = 34$ samples, or a sampling and analysis cost of $2,550.

2.9. Summary of Case Studies. Case studies 2 through 4 illustrate the multitude of related factors that must be considered when evaluating which sampling design to apply in a particular SI. When evaluating alternative sampling plans, planners may anticipate the concentration patterns likely to be present in the target population. Advanced information about these patterns can be used to design a plan that will estimate population parameters with greater accuracy and less cost than can otherwise be achieved.

2.10. Comparing On-site Data to Fixed Screening Criteria. In the data analysis phase of the SI, environmental scientists compare site data to screening values using either qualitative or quantitative statistical evaluations. The following provides a discussion of qualitative and quantitative evaluations.

2.10.1. *Qualitative Statistical Evaluations.* The EPA has developed risk-based screening criteria in the form of PRGs and RBCs. These criteria are frequently applied at the SI stage to

identify whether the site as a whole may need further attention in an RI/FS. Many screening criteria exist at both the Federal and state government level. Thus, comparisons are frequently made against the lowest of several screening criteria that can be applied to a given data set from a given location. The technical team must ensure that the criteria are being applied properly (i.e., not all screening criteria are applicable to every site), and that the implications are clear in the conclusions of the SI. For example, if site data exceed a standard developed to protect groundwater from soil leaching of contamination, but do not exceed an applicable human health standard, the team should report the results with the implications of these differences noted in the conclusions.

2.10.2. One typical qualitative method of comparing data decision limits entails the use of a spreadsheet or database. The decision limits and individual sample results are presented in a tabular format and each detected analyte concentration is compared to the corresponding screening values for that analyte. (It may be necessary to compare a single contaminant of concern to only the lowest decision limit or several different decision limits.) Table 2-1 is an example of such a spreadsheet.

2.10.3. The primary pitfall of this qualitative strategy is that the uncertainty associated with the reported results is not considered when the results are compared to the decision limits. Thus, the reported results may actually be equal to or exceed decision limits when uncertainty is taken into consideration. If this is the case, especially in the event the decision limit is exceeded, the wrong conclusion would be drawn. The ramification of an erroneous conclusion will vary, depending on the nature of the problem under investigation; nevertheless, this is an outcome that should be avoided or at least minimized.

2.10.4. Historically, environmental researchers have tended to screen analytical results into two categories—greater than the standard or less than the standard. Through advances in research and technology, three categories now exist against which analytical results can be compared: i) the reported value clearly exceeds the standard (when bias and variability are taken into account); ii) the reported value clearly does not exceed the standard; and iii) the result is inconclusive. This last conclusion is reached when the uncertainty is too large for reliable decision-making.

2.10.5. Table 2-1 illustrates how qualitative information may be used to support the decision making process when SI data are qualitatively, rather than statistically, compared to decision limits. In particular, information regarding the quality of the data, obtained in the data validation process, is used to determine whether contamination is present at concentrations greater or less than project decision limits. All applicable screening criteria are displayed in Table 2-1. For example, the “S” column reports the results of comparing each analyte concentration and the lowest screening limit. One of three codes is entered in this column for the three possible conditions identified in the preceding paragraph. An “X” is recorded if the reported values appear to be well above the decision limit, an “I” if the result is inconclusive, and a blank space if

the result appears to be well below the limit. Select results from Table 2-1 are discussed below to illustrate the nature of the screening evaluation.

2.10.5.1. Tetrachloroethene results in IRP-49 (1.2 ppb) and IRP-51 (17.08 ppb) both exceed the PRG (1.1 ppb). Although the value in IRP-49 is barely above the PRG, it reports the results as two significant figures, so we must accept its value as exceeding the PRG. However, accounting for analytical error, typically between 20 and 30% (as a conservative estimate), this result would be inconclusive. The researcher then must choose whether to conduct additional testing or accept the value of IRP-49 as an exceedance. The latter would be selected only if a conservative estimate was desired.

2.10.5.2. In IRP-49 (0.2 ppb) and IRP-51 (0.2 ppb), the reported concentration is not distinguishable from the PRG when compared on the basis of just one significant figure. Therefore, these results are inconclusive.

2.10.5.3. Several chloromethane results are marked inconclusive because of blank contamination. The only sample without blank contamination, IRP-39, was below the PRG (PRG = 1.5 ppb; IRP-39 = 0.2 ppb). The reported concentration was qualified with a J flag because it is less than the quantitation limit of 1 ppb. (The quantitation limits are not listed in Table 2-1, but were obtained from the laboratory's data package.)

2.10.5.4. For bromodichloromethane in sample IRP-48 (0.2 ppb), the reported concentration is biased low and is less than the quantitation limit of 1 ppb, so this exceedance of a PRG (0.18 ppb) is conclusive. In sample IRP-51 (0.1 ppb), the result is also biased low and is just below the PRG, so this result is also not conclusive.

2.10.5.5. For chloroform in sample IRP-39 (0.4 ppb), the reported concentration is qualified with a J flag because it is less than the quantitation limit of 1 ppb. As the reported result is quantitatively estimated, it does not reliably demonstrate that chloroform is present above the PRG.

2.10.5.6. Benzo(a)pyrene was reported in sample IRP-49 (0.278 ppb) above the PRG limit (0.0092 ppb). However, the detection limit (0.014 ppb) is above the PRG for the remaining samples. Only by achieving a lower detection limit is it possible to determine whether the non-detects are a problem. The results for benzo(a)pyrene are marked inconclusive. All of the arsenic non-detects are inconclusive based on a similar rationale.

Table 2-1.
Site Screening Data Table

		EPA MCL	Region IX PRG (1999)	IRP-39				IRP-48				IRP-49				IRP-51			
Analyte	**Units**	**Tap Water**			**L**	**V**	**S**		**L**	**V**	**S**		**L**	**V**	**S**		**L**	**V**	**S**
Organics																			
Bromodichloromethane	µg/L	—	0.18	0.1	U			0.2	J	L, s	I	0.1	U			0.1	L, s		I
Carbon Tetrachloride	µg/L	5	0.17	0.1	U			0.1	U			0.1				0.4	J	J	I
Chloroform	µg/L	—	0.16	0.4	J	J	I	0.1	U			0.1	U			0.1	U	U	
Chloromethane	µg/L	—	1.5	0.2	J	J		6.1		B	I	1.6		B	I	3.7		B	I
Methylene Chloride	µg/L	5	4.3	0.1	U			0.1	U			0.1	U			0.1	U		
Trichloroethene	µg/L	5	1.6	0.4	J	J		0.1	U			18.7			X	18.1			X
Tetrachloroethene	µg/L	5	1.1	0.1	U			0.1	U			1.2			X	17.1			X
Benzo(a)pyrene	µg/L	0.2	0.0092	0.014	U		I	0.014	U		I	0.278			X	0.014	U		I
Inorganics																			
Arsenic	mg/L	50	0.045	0.7	U		I	0.7	U		I	0.7	U			0.7	U		I
Chloride	mg/L	250	—	311			X	15.8				265			I	134.7			
Lead	mg/L	15	—	0.3	U	K		0.3	U	K		8				10			
Nickel	mg/L	—	730	590				29.0				214				198.0			
Sulfate	mg/L	250	—	44.0				5.98				41.6				21.45			
Thallium	mg/L	2	2.9	1.4				0.8	U			0.8	U			0.8	U		
Vanadium	mg/L	—	260	1.4				1.0	U			3.0				5.0			

Notes: L column contains the laboratory flags. V column contains the validation flags. S column contains screening results.

Flags: U – Not detected above reported detection limit.

B – Not detected substantially above a laboratory or field blank.

L – Biased low.

K– Biased high.

s – Surrogate failure.

J – Quantitatively estimated

Screening Codes:

X – sample concentration unequivocally exceeds the lowest screening standard.

I – sample concentration comparison to screening standard is inconclusive.

– A blank cell indicates that the sample concentration unequivocally does not exceed the lowest screening standard.

2.10.5.7. Though the reported concentration of chloride in sample IRP-49 (265 mg/L) is not qualified as estimated and exceeds the decision limit (250 mg/L), the result is marked inconclusive because the difference between the detected concentration and the decision limit is less than 5%, which is smaller than the analytical error for the test method (e.g., the error tolerance for the test method is typically 5 to 20%).

2.10.6. These results illustrate the critical importance of estimating and incorporating into decision-making knowledge of both the field and laboratory components of variance. One fundamental error is treating the reported results as conclusive when in fact they are not. The values represented in this table are measurements, and measurements contain bias and variability that must be accounted for in decision-making. (See EM 200-1-10 for additional guidance on the data review strategies that were primarily used to qualify the results in Table 2-1.)

2.11. Quantitative Statistical Evaluations. When the results of the qualitative statistical evaluations are inconclusive, further investigation is required. DQOs must be revised so that the parameter of interest is no longer a single datum per location. Instead, multiple samples are collected for those uncertain locations and the resulting distribution of values is compared to the decision limit using quantitative statistical tests. The results would typically be statistically compared to decision limits using *one-sample tests*[*] for central tendency, as discussed below.

2.11.1. All statistical tests require the user to make certain assumptions about the data to perform the statistical test. The user must demonstrate that the underlying assumptions for a particular statistical test are reasonable before doing the test. With respect to these underlying assumptions, statistical tests can be roughly categorized as either *parametric*[†] or *non-parametric*.[†] When non-parametric tests are conducted, data sets are required to satisfy fewer assumptions than for the corresponding *parametric tests*.[†] In particular, a parametric statistical test assumes a specific *distribution*[†] for the data (i.e., the entire population is described by some specific mathematical function), such as the bell-shaped curve for the *normal distribution*[‡]. Statistical plots of actual measured sample concentrations must be substantively consistent with the corresponding plots generated using the theoretical functional relationship. Tests that require normal or log normal distributions are most commonly used. (A data set is log normal if, when the log of each datum is calculated, the resulting set of values is normally distributed.) Common graphical methods (i.e., plots) are presented in Appendix J. In addition, an overview of the evaluation of distribution assumptions is presented in Section III of Chapter 3.

2.11.2. It should also be noted that parametric tests become problematic, and may not be possible to perform, when the data sets contain a significant number of *censored*[§] values (i.e., analyte concentrations reported as non-detects). However, as described in Appendix H, it may be

[*] Appendix L.
[†] Appendices H and I.
[‡] Appendices E, F, and J.
[§] Appendix H.

possible to use the *Poisson distribution*[*] for highly censored data. Parametric tests are also problematic when there are outliers. The possibility of *outliers*[†] should be considered in every analysis.

2.11.3. Non-parametric tests do not assume a specific functional relationship for the data distribution. These tests tend to be less sensitive to outliers and non-detects than parametric tests. Although non-parametric tests are more applicable relative to parametric tests, non-parametric tests tend to be less statistically *powerful*[‡] than parametric tests. In essence, this means that more samples must be collected for a non-parametric test relative to the corresponding parametric test to make decisions at the same level of confidence.

2.11.4. Background concentrations of naturally occurring and anthropogenically derived compounds are also possible screening criteria. However, there are few instances in which such background levels are available at the SI stage. Sometimes a "site-wide" statistical background study has been done. If such a study is available, *two-sample statistical tests*[§] would be used to compare the study area data set with the "site-wide" background data set. (As the name implies, a two-sample statistical test is predominantly a statistical evaluation to compare two separate sets of data.) Because an RI often includes specific sampling for background, the determination of background levels and their usefulness is described in Chapter 3. If the SI is the first sampling event for a site, there is a low probability that site-specific background sample data exist.

[*] Appendices E, G, and H.
[†] Appendix I.
[‡] Appendix O.
[§] Appendix M.