CHAPTER 3
Remedial Investigation/Feasibility Study (RI/FS)

3.1. Introduction. If, based on the PA/SI, a site warrants listing on the National Priorities List (NPL), an RI/FS is performed at the site.

3.1.1. The RI is the stage in the CERCLA process for collecting data to do the following.

3.1.1.1. Characterize site conditions (e.g., thickness of unsaturated soil [vadose zone], depth to groundwater, vegetative cover, background conditions).

3.1.1.2. Determine the types, conditions, and distribution of the waste contamination in affected media.

3.1.1.3. Assess risk to human health and the environment.

3.1.1.4. Conduct treatability tests to evaluate the potential performance and cost of the treatment technologies that are under consideration.

3.1.2. The FS is the stage for the development, screening, and detailed evaluation of remedial actions.

3.1.3. The RI and FS are intimately linked. Data from the RI influence the development of remedial alternatives in the FS, which in turn affect the data needs and scope of treatability studies and additional field investigations. This phased approach encourages the planning team to continually plan the site characterization effort, which minimizes the collection of unnecessary data and maximizes data quality.

3.1.4. As in the SI phase, the initial statistical elements in the RI process involve the development of DQOs. The statistical evaluations used for the RI typically include those performed for the SI. For example, as in the SI, site data are often statistically compared to some set of fixed decision limits and upper confidence limits are often established (as discussed in Chapter 2). In general, the statistical evaluations are more common for RIs than SIs, and the statistical analysis tends to be more comprehensive. In part, this is because typically data coverage is greater and the RI data quality objectives are more robust. For example, while the SI predominantly focuses on statistical evaluations to resolve the presence or absence of contamination, the RI reaches for a determination of the extent of contamination. Critical to the onset of an RI is the identification of Applicable or Relevant and Appropriate Requirements (ARARs), which, in turn, may influence the identification of areas requiring remediation. Both sampling strategy and extent of contamination are influenced by the selection of ARARs. ARARs help identify the best analytical procedures needed to reach decision limits. This aspect of DQOs is addressed in Appendix C.

Section I
Site Characterization

3.2. Introduction. The first two objectives of the RI (subparagraphs 3.1.1.1 to 3.1.1.4) are combined for discussion in this Paragraph. The process of site characterization is linked to the procedures described in Section II of Chapter 2, where sampling distribution design was discussed. In the RI stage, sample design is likely to be influenced by SI data. In turn, these SI results affect the statistical methods at the planner's disposal for collection of site data.

3.2.1. When scoping for the SI, project planners have expectations about the probable location and nature of contamination. By the time a site reaches the RI, some usable information is usually available. In particular, if a contaminant was identified in the SI, planners may have an idea of the mean and standard deviation of contaminant concentrations. These initial estimates assist in devising a statistical sampling design at the RI stage. Two examples of using site data to support sampling design are presented in this Paragraph. These are "hot spot" sampling and geostatistical sampling, the fundamentals of which are presented in Appendices C, J, and Q.

3.2.2. A "hot-spot" typically refers to a localized area of high concentration, but is often otherwise poorly defined (e.g., criteria for the size and concentration of hot spots are often arbitrary or not specified). Hot-spots are not uncommon at sites where waste was released in an isolated region, perhaps during a spill. In addition, hot-spots may occur within broader regions with low, but detectable, levels of contamination. One example of this may be when an area was used to process waste disposal over some time and, at times when a shop or operation was cleaning house, a high concentration of waste would be deposited. However, sample concentrations that exceed a regulatory threshold or other decision limit should not be considered to be hot-spots if these concentrations appear to be randomly distributed and will not necessarily be of concern if they represent a small portion of study area and contain a small contaminant mass.

3.2.3. Case study 1 presents an RI application of the hot-spot identification method discussed in Appendix C.

3.2.4. In this instance, professional judgment led to the determination of the size and shape of the hot-spot. The reader is urged to vary S and L to identify the sensitivity of hot-spot sampling grids to the assumptions.

3.2.5. As stated previously, there is typically some knowledge of contaminant distribution at a site by the time an RI begins. Geostatistics allow an investigator to extrapolate (and interpolate) what is known in one location to other nearby related locations. Its application relies on the fact that, given a known concentration at one location, an adjacent location is likely to have a similar concentration. The greater the distance from the known concentration, the greater uncertainty there is in predicting a concentration at an unsampled location. This situation can be de-

scribed as a spatial correlation, because correlations are related to how close samples are to one another. Geostatistical methods are described in detail in Appendices J and Q.

3.2.6. Case Study 2 illustrates the use of geostatistics for reducing uncertainty in a project. Although geostatistical techniques are more common for RIs than SIs, they may also be used for SIs if sufficient site data are available.

3.2.7. One of the major RI objectives is identifying the distribution of contamination at a site. As useful as geostatistics are in helping with sampling design, they may also be used in interpreting sample data. The geostatistical method known as kriging (Appendix J) is an effective method for interpolating site concentration data under conditions where spatial correlation exists. Kriging is a weighted-moving-average interpolation method. The USEPA developed a two-dimensional kriging package, which is useful in providing a fundamental introduction to the technique (Geo-EAS; EPA/600/4-88/033). Kriging as a method of contouring is described in several readily available texts, and typically requires the use of commercially available computer software with kriging options for contouring (e.g., Surfer, EVS).

3.3. Case Study 1—Hot-Spot Identification. The project team attempted to locate a hot-spot resulting from an uncontrolled water release within a larger storage area. The total storage area was approximately 150 by 200 feet. Because the suspected waste was spilled as a liquid, the hot-spot was assumed to be approximately circular. A best estimate of the diameter was approximately 20 feet. The method proceeded in steps as follows:

3.3.1. A circular hot-spot means S equals 1.

3.3.2. The radius of the target spot is 10 feet.

3.3.3. The team assigns a value of 0.1 to the acceptable risk of not finding the hot-spot.

3.3.4. Using S and β, refer to Table D-1 (or nomographs presented in Gilbert, 1987) to determine that L/G is 0.55 for a square grid and 0.50 for a triangular grid.

3.3.5. Using the relationship L/G and the assumed radius of 10 feet, we see that square grid spacing is 18 feet and triangular grid spacing is 20 feet (values are rounded to the nearest foot to reflect the significant figures).

3.3.6. One sample will be placed at each grid node in the storage area, so that a square grid requires 88 samples and a triangular grid requires 75 samples.

3.4. Case Study 2—Using Geostatistics in Project Planning to Reduce Uncertainty and Cost. At a site in the Midwest, project planners were asked to assess a site potentially contami-

nated with lead at levels exceeding risk-based limits. A SI was conducted using a grid system over areas that were suspected of being contaminated based on historical information.

3.4.1. The project team identified lead concentrations in soil exceeding threshold values in various areas of the site (red circles in Figure 3-1). They were required to move on to an RI/FS to more fully characterize the nature and extent of contamination and develop preliminary estimates of cost for a removal action. Initially, the team intended to collect numerous additional samples on a grid (green circles in Figure 3-1) to more fully delineate the extent of contamination. However, the project geologist suggested the use of geostatistics as a means of reducing the number of samples without increasing uncertainty.

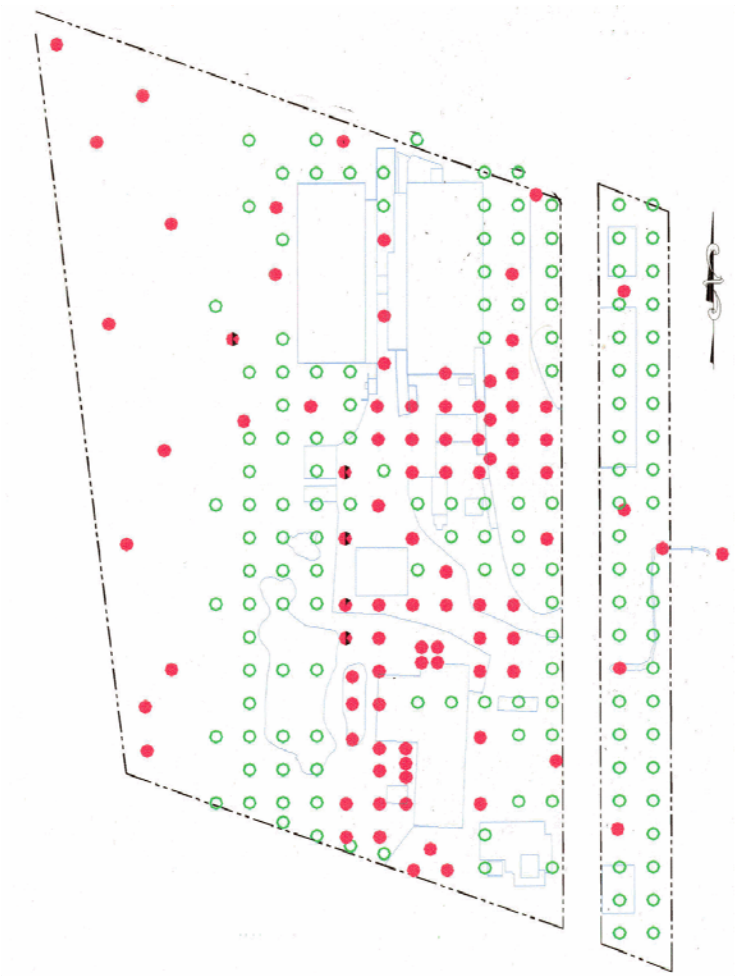

Figure 3-1. Initial sampling grid and proposed new samples.

3.4.2. Geostatistics can predict both the concentration and the uncertainty for an unsampled portion of the study area. In essence, spatial correlations for contaminant concentrations es-

tablished from the existing data set are used to "extrapolate" sample concentrations and uncertainty for other portions of the study area. Consequently, the team was able to use a geostatistical evaluation to assess the value of collecting additional samples at any given location in the grid. Simply put, the team recognized that in any sampling and analysis system there will be bias and variability, and that estimates of that bias and variability could be made using the existing data. Thus, at any location where the estimate of uncertainty from the geostatistical prediction was less than the uncertainty from sampling and analysis, the team reasoned that there was no value in collecting additional samples.

3.4.3. The final sampling plan required the addition of only seven new sampling points (shown as black circles in Figure 3-2) with associated cost savings of over $12,000.

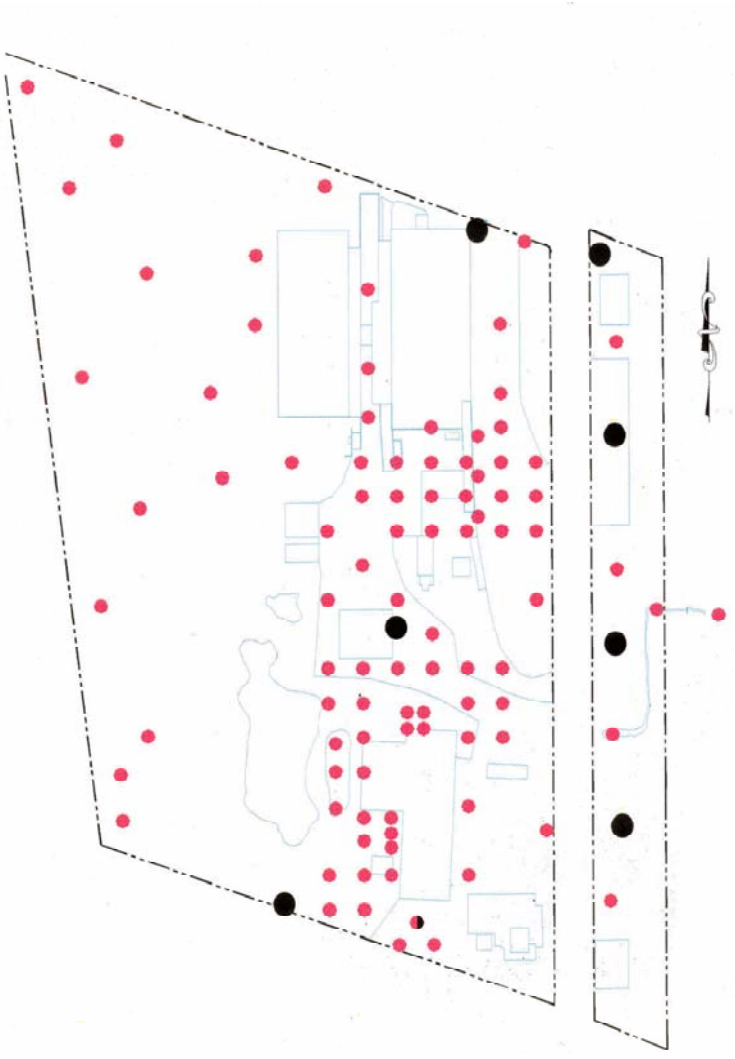

Figure 3-2. Samples required after geostatistical analysis.

Section II
Background Comparisons

3.5. Introduction. Not all chemicals detected at hazardous waste sites originate from site-related activities; for example, metals in soil and groundwater are often present because of natural geological conditions. Similarly, anthropogenic activities unrelated to a site frequently contribute certain organic chemicals (e.g., polycyclic aromatic hydrocarbons [PAHs] or pesticides derived from urban or agricultural sources; EPA SOW No. 788). If site sample concentrations for a specific compound are similar to or lower than background concentrations[*], there may be no need to consider potential remedial actions with respect to that compound. This determination can be quantitatively defended by use of statistical comparison methods.

3.5.1. The project team should determine the background sampling locations and parameters during the planning stages of the RI. Separating and identifying background sample locations from portions of the study area that have been potentially affected by waste handling activities is an example of stratification. The critical factor distinguishing a background sample from the site lies in understanding where contaminated areas end and natural conditions begin. Such samples may be located upwind, upstream, or upgradient from the waste site. Background data should be drawn from media that physically represent the study area; they should be from the same soil type or geological deposit, same type of surface water system (for example, freshwater versus saltwater; wet season versus dry season), or from the same aquifer as the site data. It is also critical to collect the background samples in substantively the same manner that the site samples are collected (same analytical method, volume of sample, etc). The sampling design and analytical methodology for the background and the site study areas must be similar. For example, erroneous conclusions can result if judgmental sampling is done for the site study area but random sampling is done for the background study area.

3.5.2. Background locations should be in a nearby portion of the region unaffected by site activities. As a caveat, site planners should be skeptical if regulators prefer to limit background sampling to only pristine areas; doing so will potentially result in erroneously concluding that the study area has been adversely impacted by site-related waste handling activities.

3.6. Does Background Soil Differ From Site Soil? The USEPA has developed guidance for addressing whether site soil characteristics differ from background (EPA/540-R-01-003 and EPA/540/S-96/500). The guidance EPA/540-R-01-003 emphasizes the formulation of DQOs in devising background sampling design and subsequent site to background testing. The focus of the cited guidance is only to determine whether site and background soil chemistry differ. It does not establish comparison standards, or levels of background that may replace unnaturally low risk-based clean-up goals.

[*] Background does not mean pristine or unaffected by human activity, especially at sites in heavily industrialized areas.

3.6.1. Fundamentally, the USEPA guidance (EPA/540-R-01-003) identifies two forms of background testing:

3.6.1.1. *Background Test Form 1*. Tests the null hypothesis that the mean contaminant concentration in samples from the site waste handling area is less than or equal to the mean concentration in background areas.

3.6.1.2. *Background Test Form 2*. Tests the null hypothesis that the mean contaminant concentration in samples from the site waste handling area exceeds the mean concentration in background areas by more than a specified margin (e.g., by 50 ppm).

3.6.2. Before continuing with this approach, investigators need to be certain that these tests are applied to random sample data sets collected from both the site and background locations. Typically, site sampling may have a component of judgmental sampling, meaning samples were biased to expected contaminated areas of a site. In such cases, the background testing cannot be applied.

3.6.3. The project planning team should establish which form of background testing will be applied at the onset of the RI planning process. In addition, the planning team needs to establish the levels of acceptable levels of error in the decision-making. This will differ from site to site, and will depend on the desires of the project planning team members.

3.6.4. The USEPA guidance also provides examples for the application of test methods that may be applied to the background test forms (EPA/540-R-01-003; Table 3-1). These are:

3.6.4.1. *Descriptive Summary Statistics*. These (e.g., mean, median, standard deviation, variance, percentiles—see Appendix D) may be used as a preliminary screening tool for comparison with site history and land use activities in the establishment of background. EPA considers these "simple and straightforward [but having low] statistical rigor."

3.6.4.2. *Simple Comparisons*. These (i.e., greater than maximum) may be used with very small data sets. This approach is not recommended.

3.6.4.3. *Parametric Tests*. These (e.g., Student *t*-test–see Appendix F) may be used if a larger number of data points is available ($n > 25$). EPA states that parametric tests require approximate normality of the estimated means and recommends that, for smaller data sets, investigators examine data for normality or lognormality in distribution. EPA considers this application statistically robust enough to be used frequently in parametric data analysis.

3.6.4.4. *Nonparametric Tests*. These (e.g., Wilcoxon Rank Sum Test—see Appendix M) may be used when data are not normally distributed, as rank-ordered tests make no assumption

on distribution. Again, EPA considers this approach statistically robust and to be used frequently in background estimation.

3.6.5. The list of methods is not complete, but, by reviewing the appropriate Appendix, users of this Manual may identify the most appropriate statistical method for site application. USEPA guidance leans heavily toward parametric and nonparametric tests, which in turn rely on establishing whether data are normal or lognormal (see Appendix F).

3.6.6. The U.S. Department of the Navy (DON) also developed statistical guidance for evaluating background in soils (UG-2049-ENV). Like the USEPA method, the guidance suggests comparative methods for testing whether site data differ from background. However, DON guidance is unique, in part, because it also relies on geochemical relationships. UG-2049-ENV provides guidance for evaluating the geology of the site and the geochemical characteristics of site soils as they relate to background analyses. The procedures outlined in UG-2049-ENV can be quite useful for USACE projects and are recommended as a resource for additional reading.

3.6.7. This "geochemical method" is often used when reference area data are not available. The method may be used to extract background concentration ranges by evaluating correlated background chemicals using on-site data only (i.e., no background area need be sampled). The key concept is that if the site has not been affected by a release, then only one population exists at a site; if a release has affected the site, then overlapping of different population characteristics would be evident in the data.

3.7. Simple Background Comparison. Investigators are more likely to rely on regional background at the SI stage than the RI. As the text below states, site-specific background is more desirable, but SI project budgets rarely allow for a full background study and such regional comparisons are still useful. Background concentrations are typically not known prior to RI activities, and sampling for background should be scoped in the planning stages of the RI. In some instances, background criteria are available as regulatory limits, as Case Study 3 illustrates. (Although the case study could also apply in an SI [Chapter 2], it is presented here to illustrate the concepts that arise for background comparisons all in one section of this document.)

3.8. Case Study 3—Comparison to Regional Background. Site-specific background concentrations are typically not known prior to RI activities, and sampling for background should be scoped in the planning stages of the RI. In some instances, regional background values may be compared to site data.

3.8.1. Texas has established soil background levels that can be used in the screening process if site-specific background levels are not available. Soil data from one site proposed for redevelopment were compared to Texas background levels. Texas regulation states that if the maximum concentration of the chemical under investigation does not exceed the Texas soil background level, then that chemical is not of concern. The site analytical data were reviewed for

quality and applicability. Based on the review, the project team was satisfied that the site analytical data were of sufficient quality for use in evaluating the site. The soil analytical data (in mg/kg) for chromium were:

6.17	4.31	4.38	6.07	5.68
2.86	5.08	4.98	2.22	15.30
4.75	3.56	4.48	3.46	2.63

3.8.2. The maximum concentration for chromium at the site is 15.30 mg/kg. The Texas soil background level for soil is 30 mg/kg. Therefore, chromium would not be a chemical of concern at the site.

3.8.3. As indicated in the USEPA guidance, such a comparison lacks statistical rigor, but is useful for guiding the project planners in the next phase of investigation.

3.8.4. At this stage, the comparison to regional background is merely sufficient to proceed to additional phases of site chromium evaluation.

3.9. Parametric and Nonparametric Tests. In the preceding case study, the regulatory community established background concentrations. It is far more desirable for local background levels to be assessed and applied. Differences related to sample medium, sampling method, or analytical method are less likely to arise in site-specific background data than regional background data. However, the project must be budgeted for a sufficient number of samples to characterize site-specific background conditions; a large number of samples may be required to characterize heterogeneous background media. If the regional background data (e.g., the background data from a very limited site-specific background study) are shown to be statistically different from a waste site, it may also be attributable to differences in water quality or soil types between the site and the location where the regional background data were collected, and not necessarily related to a waste release. Therefore, a thorough evaluation of local background conditions is preferred to the use of regional background levels.

3.9.1. Instructions and guidance for selecting analytical procedures as part of DQOs should be applied to the background data set with the eventual uses of background data in mind. For statistical comparison, background measurements need to be random. In addition, the power of statistical comparison may be greater if the background results are normally or lognormally distributed. Although the distribution of background measurements cannot be guaranteed, either random or systematic sampling of background should be a component of the sampling plan. (Note that given spatial correlation, systematic samples spaced closer than the geostatistical range may not be independent. Sampling methods are addressed in Appendix C.) Once a set of background samples have been collected, comparison methods are applied using the statistical procedures addressed in Appendix M or N.

3.9.2. A *random sampling*[*] design is typically used to characterize the background study area. *Two-sample statistical tests*[*] are then typically used to compare the site data set to the background data set. Two-sample tests, described in Appendix M, are summarized in Table 3-1.

3.9.3. An example of determining COPCs using background population tests is presented in case study 4.

Table 3-1.
Background Population Comparison

Percent Detections in Site Data	Percent Detections in Background Data	Test
0–100	0	No comparison
> 0–100	< 10	Poisson UTL
10–50	10- 50	Test proportions
> 50	> 50	Mann-Whitney test,
85–100	85–100	Student's *t* test* or Mann-Whitney test
*Student's *t* test should be used if the distributions in the site and background data sets are the same; otherwise, the Mann-Whitney test should be used.		

3.10. Case Study 4—Establishing and Comparing Background Concentrations to On-site Data. At a military installation in Utah, samples were collected for metals in soil—seven on site and four at background locations. This case study focuses on chromium. The chromium results were as follows (mg/kg):

SS01	SS02	SS03	SS04	SS05	SS06	SS07	BKG1	BKG2	BKG3	BKG4
4.3	2.7	2.2	3.2	<1	3.6	2.4	1.6	1.8	2.6	1.6

3.10.1. Because the site data had an 85% detection rate, one-half the reporting limit was substituted for each non-detect for the statistical calculations.

3.10.2. Both background and site data were determined to be normally distributed at a 90% confidence level. An *F-test* was used to compare the variance of the background data set to the variance of the site data set. The result of the *F*-test indicated that the variances are equal.

3.10.3. Thus, a *two-sample t-test* (with equal variances) was used to compare the background and on-site data sets. At the 95% confidence level, the calculated $p = 0.172$. Based on this evidence, a statistical difference between background and on-site data could not be demonstrated at the 95% level of confidence; thus, no further action with respect to chromium was required. Note that, for this simple example, the conclusion of "no further action" is drawn because a statistical difference was not obtained. The power of the test is normally calculated when the

[*] Appendix C.

null hypothesis is not rejected. Additional investigation would be required if the power was not adequate.

3.11. Upper Tolerance Limits. *Upper tolerance limits*[*] (UTLs) are sometimes used to determine whether site concentrations are elevated relative to background concentrations. The UTL defines a threshold value for the background data set. (More accurately, it is an upper confidence limit for some percentile of the background data.) Individual site contaminant concentrations are compared to this value. Study area detections that are greater than the background UTL are considered to be indicative of contamination from site-related waste handling activities. Tolerance limits are used in this manner in the USEPA guidance for the statistical treatment of groundwater monitoring data (EPA 530-SW-89-026, EPA 9285.7-09A). However, this approach must be used with caution. In particular, it is often erroneously concluded that site-related contamination exists if a single detection exceeds the UTL. For example, the "95% UTL" is typically used to evaluate site contamination relative to background. If the background and site concentrations are not different from one another, we will be 95% confident that *at least* 95% of all site measurements will fall below the *95% UTL with coverage of 95*%. (For brevity, this is often referred to simply as the "95% UTL.") Therefore, we would expect a small percentage of site measurements to exceed the UTL, even when overall site contamination is not elevated relative to background. When a large number of samples are taken, we should not definitively conclude that a small number of detections greater than the UTL necessarily indicate site-related contamination.

3.11.1. Furthermore, regulators have criticized the use of UTLs to compare site to background contamination because UTLs do not minimize false negatives but, rather, minimize false positives. In other words, if many detected study area concentrations were greater than the background UTL, this would constitute strong evidence of site-related contamination. This scenario would be unlikely if the site and background concentrations were similar. Alternatively stated, the probability of a false positive—erroneously concluding that the site is contaminated relative to background—would be low. However, if detected site concentrations were less than the UTL, strictly speaking; *no conclusion would be possible*. This would not be sufficient to demonstrate the absence of site contamination relative to background. If we were to conclude the absence of site-related contamination using the UTL, false negatives could result (i.e., erroneously concluding that site concentrations are not elevated relative to background concentrations).

3.11.2. Because of the problems with tolerance intervals discussed above, two-sample statistical tests are usually preferred (and are typically more appropriate) to compare site and background data sets. It is recommended that UTLs be used only when two-sample tests are not practical (or when the primarily objectives is to demonstrate that site contamination is elevated relative to background contamination). For example, a two-sample statistical test cannot be performed when the site data set is extremely small (when only one or two samples are available for the study area). If a large data set was available for the background study area (e.g., because a

[*] Appendices G and K.

"site wide" background study had been done for a prior investigation), then the study area results could be compared to the background UTL.

3.11.3. The UTL background comparison methods are discussed Appendix K. These methods are summarized in Table 3-2

3.11.4. There are parametric UTLs and non-parametric UTLs. The parametric UTL require the data to follow a specified distribution such as a *normal* or *lognormal* distribution. (Distribution tests are addressed in Appendices F and J.) As shown in the table above, the proportion of non-detects must be taken into account when selecting an appropriate UTL. (UTLs that rely upon the normality assumption cannot be calculated when a large portion of the data are reported as non-detect.) The nonparametric UTL represents a high-end value in the distribution. The following case study illustrates an example of calculating background UTLs for metals.

Table 3-2.
Background Comparison to Evaluate the Extent of Contamination

Percent Detections in Background Data	Type of UTL Calculated
0	No UTL calculated
< 10	Poisson UTL
10–85	Nonparametric UTL
≥ 85 (normal or lognormal distribution)	Parametric UTL

3.12. Case Study 5—Calculating Background UTLs for Metals. At a site in Utah, 56 soil samples were collected across a very large area to determine background concentrations for metals.

3.12.1. Chromium was detected above the detection limit in every sample, so there was no need to substitute for censored values. Manganese was not detected in one sample, and the geochemist elected to substitute one-half the detection limit for the censored value in that sample.

3.12.2. The chromium data were normally distributed and the manganese data were *lognormally distributed.*[*] Refer to Appendices D, E, and I for a review of these concepts.

3.12.3. For chromium, the 95% UTL was calculated from the sample results using the formula:

$$95\%\ \text{UTL} = \bar{x} + ks \ .$$

3.12.4. For 56 samples, k equals 2.032. Chromium results for background had a mean ($\bar{x}$) of 12.7 mg/kg and standard deviation of 5.1 mg/kg, so the UTL was 23.0 mg/kg. For manganese,

[*] The *Shapiro-Wilk test* (Paragraph F-3) was used to test for normality at the 95% level of confidence.

the log of each sample result was taken prior to the calculation of the UTL. (The individual concentrations are not shown.) For the set of log-transformed results, the sample mean and standard deviation were 5.41 and 0.75, respectively. The log UTL for manganese was 6.93 (using the above equation). All comparisons for manganese should occur in "log space" (that is the logarithm of the site manganese maximum would be compared to 6.93). (Alternatively, a *minimum variance unbiased estimator* of the manganese background concentration could be calculated using the methods described in Appendix E).

3.13. Extended Background Example. This paragraph illustrates the concepts of distributional assumptions presented in Appendix J through a case study.

3.13.1. Suppose surface soil samples (from 0 to 5 feet below ground surface) have been collected at Site A and a background location to evaluate chromium concentrations on site. Table 3-3 presents the analytical results from samples collected at the site and background areas. All chromium concentrations were detected so no proxy concentrations are needed to evaluate the data.

3.13.2. Further, suppose the objectives of this data evaluation are to identify whether chromium surface soil concentrations on site:

3.13.2.1. Exceed regulatory threshold levels.

3.13.2.2. Exceed background concentrations, on the average.

3.13.3. Several statistical tests can be used to make such comparisons. A "one-sample" test can be used to compare the mean site chromium concentration to regulatory risk-based levels (Appendix L). A "two-sample" test can be used to compare the mean concentration of chromium at the site to the mean background concentration of chromium (Appendix M). A background value, such as a UTL, can be estimated for comparisons to individual site concentrations to identify if any one sample has a concentration higher than background. However, before any statistical tests can be done, distributional assumptions must be evaluated for each population (site and background) of data to determine which statistical test is most appropriate. The distributions are evaluated for normality (or log normality) using statistical tests and graphical plots.

3.13.4. Graphical displays are the first approach taken to evaluate the distribution of the data (Appendix J). Histograms, box-and-whiskers plots, and probability plots are all useful in identifying how data are distributed and answering questions such as—are the data symmetrical, what is the range of concentrations, are there any outliers that may unduly influence future distributional tests, do the data seem to follow a normal distribution, and so on. Histograms, box-and-whisker plots, and probability plots for the site and background data are provided in Figures 3-3 and 3-4, respectively.

Table 3-3.
Analytical Results for Chromium at Site A and Background Locations

Site A Sample Location	Top Depth of Sample	Bottom Depth of Sample	Chromium Concentration (mg/kg)	Background Sample Location	Top Depth of Sample	Bottom Depth of Sample	Chromium Concentration (mg/kg)
SB01	1	2	4.76	BG01	1	2	4.99
SB01	4	5	4.42	BG01	4	5	4.35
SB02	1	2	4.68	BG02	1	2	4.61
SB02	4	5	4.82	BG02	4	5	4.83
SB03	1	2	4.36	BG03	1	2	3.92
SB03	4	5	4.37	BG03	4	5	5.09
SB04	1	2	4.09	BG04	1	2	5.19
SB04	4	5	4.14	BG04	4	5	4.54
SB05	1	2	4.78	BG05	1	2	5.49
SB05	4	5	4.94	BG05	4	5	4.3
SB06	1	2	3.35	BG06	1	2	5.67
SB06	4	5	3.08	BG06	4	5	4.16
SB07	1	2	10.1	BG07	0.5	1	5.41
SB07	4	5	18.5	BG07	2	2.5	4.98
SB08	1	2	10.6	BG08	1	2	5.64
SB08	4	5	4.87	BG08	4	5	4.98
SB09	1	2	10.3				
SB09	4	5	5.51				
SB10	1	2	6.4				
SB10	4	5	4.13				
SB11	1	2	4.96				
SB11	4	5	4.96				
SB12	1	2	4.91				
SB12	4	5	4.89				

3.13.5. These plots have been developed on the basis of the original data and the natural-log transformed data, as it is common that environmental data follow either a normal or log-normal distribution. Other less common transformations, such as the square root or inverse sine transformation, are *not* applicable in this case study because:

3.13.5.1. Chromium concentrations are continuous (values can be any number within a range of concentrations).

3.13.5.2. Detected chromium concentrations are not rare events to warrant review of the Poisson distribution.

3.13.5.3. Chromium concentrations are not binomially distributed.

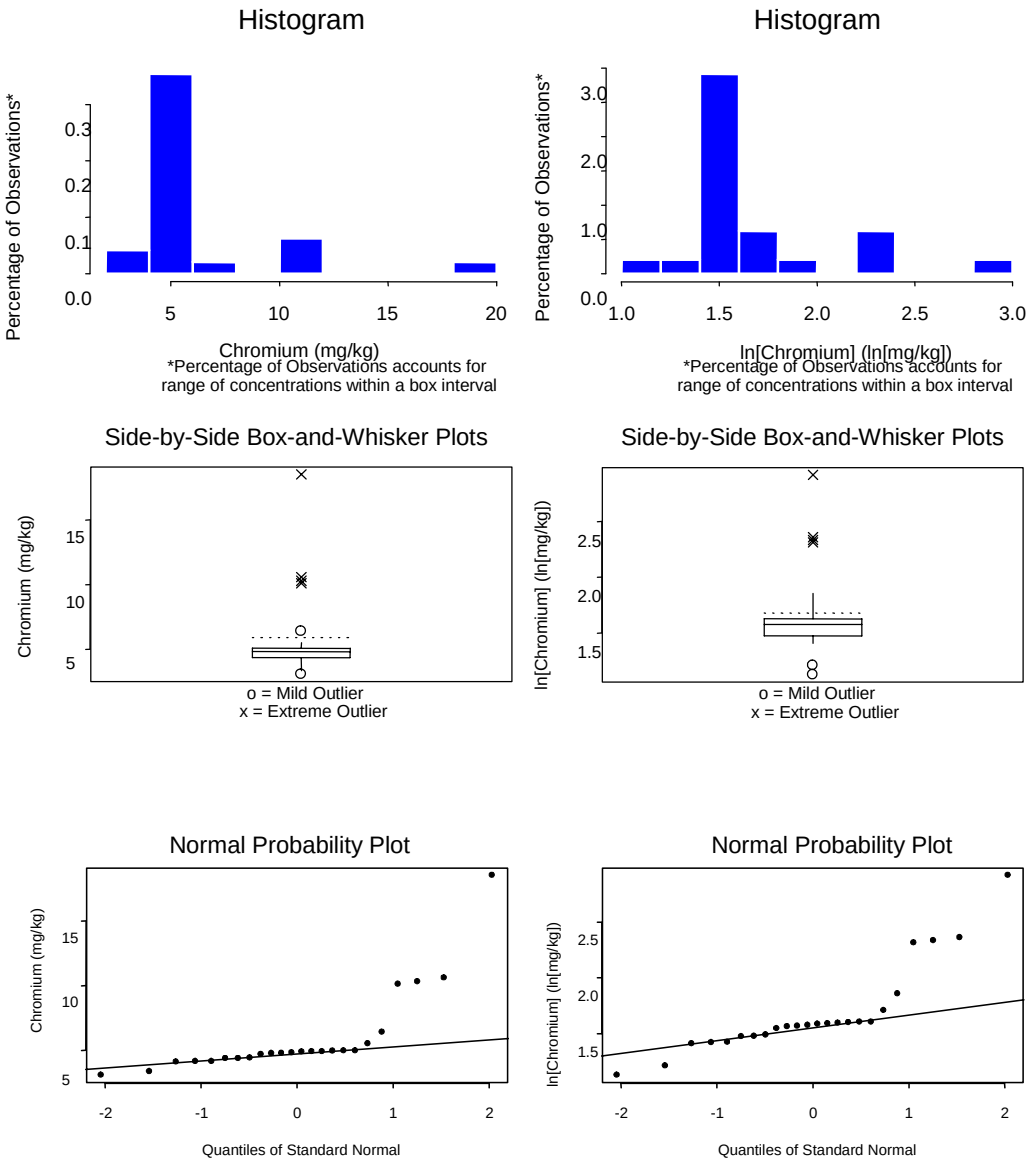

Figure 3-3. Chromium in Site A.

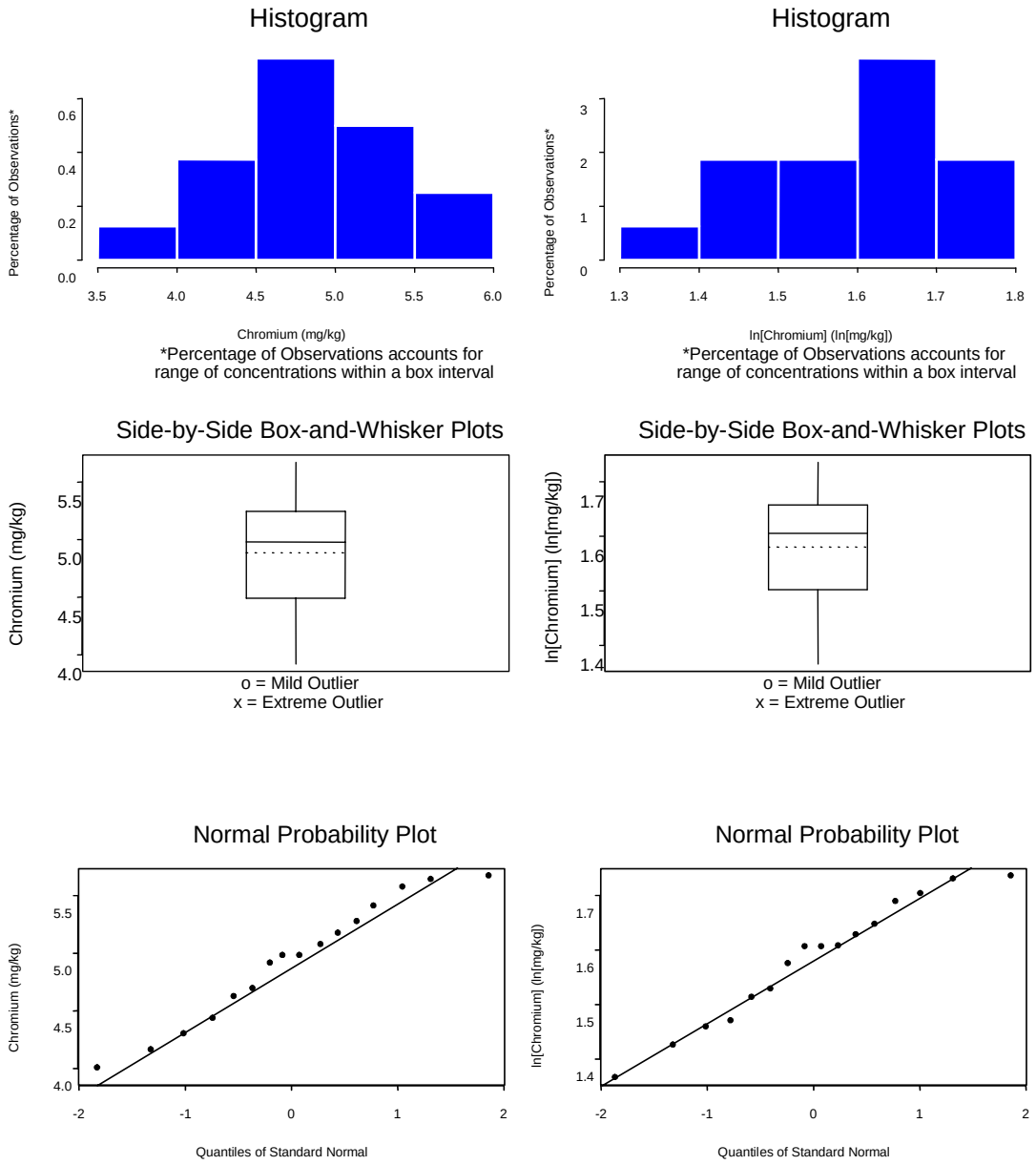

Figure 3-4. Chromium in background.

3.13.6. Based on just the plots in Figure 3-3, chromium at Site A does not appear to have a normal or lognormal distribution. The histograms for the original data and log-transformed data are not symmetrical, but are skewed. This is confirmed in the box-and-whiskers plots because the mean (the dotted line) is larger than the median (the solid line within the box) and the mean is even larger than the 75th percentile (the top part of the box). (If the data were normal, the mean

would be equal to the median.) As the mean is greater than the 75th percentile, this suggests that the mean is influenced by several considerably large concentrations. Outliers (each of point represented by an "X") predominantly occur only in the upper portion (the top) of the box plots. Lastly, as the normal probability plots for the original data and log-transformed data are not linear, this gives additional evidence that the data are not normal or lognormal.

3.13.7. The chromium data distributions possess heavier right tails relative to a normal distribution. Note the extreme deviation from linearity (Appendix F) at the right-hand side of each normal probability plot (appearing as a series of points above the straight line). The superimposed line on the normal probability plots illustrates the line that concentrations follow when data are normally or lognormally distributed. This line is related to Filliben's statistic in the sense that it provides a standard to compare the linearity of sample results. For these normal probability plots associated with Site A, it is apparent that the data do not follow a normal or lognormal distribution.

3.13.8. The plots in Figure 3-4 show evidence that chromium for the background data set appears to follow a normal or a lognormal distribution. The histogram for the original data seems to be symmetrical, though the histogram for the log-transformed data is not as symmetrical. However, histograms can be misleading if the boxes (i.e., concentration intervals) are too large or too small; therefore, another type of plot, preferably a normal probability plot, should be constructed to determine whether the data are normally (or lognormally) distributed.

3.13.9. One of the most powerful statistical methods for testing normality is the *Shapiro-Wilk*[*] test. Because the site data set has 24 sample results and the background data set has 16 sample results, this test would be appropriate for evaluating normality and lognormality for both the site and background data sets. The result of the Shapiro-Wilk test is presented in Table 3-4 for chromium at Site A and background based on the original data and log-transformed data. The Shapiro-Wilk test results in either a calculated value of the statistic W or the value p. There is acceptably strong evidence that the data set is not normal when either W or p is small relative to the corresponding acceptance limit for W or p.

3.13.10. For Site A, results of the Shapiro-Wilk test show evidence that the data *do not* follow a normal or lognormal distribution (i.e., since the calculated value of W is smaller than $W_{0.01}$, or equivalently, $p < 0.01$, there is less than a 1% chance that the data set is normal, or equivalently stated, there is at least a 99% confidence that the data are not normal). However, for background the results of the Shapiro-Wilk test suggest that the data seem to follow both a normal and lognormal distribution. It should be noted that there is more evidence that background data are normally distributed rather than lognormally distributed, because the value of W and the associated value of p are higher for the original data than for the log-transformed data.

[*] Appendix F.

3.13.11. The *coefficient of variation*[*] (*CV*) was estimated for each data set, and is provided in Table 3-4. A *CV* greater than 1 suggests a departure from normality. However, the evaluation of the *CV* is not as reliable as quantitative statistical tests for normality, such as the Shapiro-Wilk test. The coefficient of variation is useful only for identifying obvious departures from normality when *CV* is much greater than 1. Because the sample *CV*s for the site and background data sets based on the original data and the log-transformed data all are less than 1 (as discussed in Appendix F), one cannot conclude the data can be modeled by a normal distribution. Therefore, for these data sets, the *CV* does not provide any useful additional information.

3.13.12. Similarly, to illustrate the relative reliability of various distributional test methods, the *Studentized range test** was also performed on the data sets. The results of this test (Table 3-5) indicate that the Site A and background data sets follow normal and lognormal distributions. The range test failed to identify the lack of normality for Site A data. This happened because the data distribution for Site A is asymmetrical and this test does not perform well for asymmetrical distributions. However, according to the test, the background data follow a normal and lognormal distribution. Therefore, the Studentized range test for the background data set is consistent with the Shapiro-Wilk test, the coefficient of variation test, and the graphical plots (e.g., the normal probability and box plots).

3.13.13. Similarly, to illustrate the relative reliability of various distributional test methods, the *Studentized range test** was also performed on the data sets. The results of this test (Table 3-5) indicate that the Site A and background data sets follow normal and lognormal distributions. The range test failed to identify the lack of normality for Site A data. This occurred because the data distribution for Site A is asymmetrical and this test does not perform well for asymmetric distributions. However, according to the test, the background data follow a normal and lognormal distribution. Therefore, the Studentized range test for the background data set is consistent with the Shapiro-Wilk test, the coefficient of variation test, and the graphical plots (e.g., the normal probability and box plots).

3.13.14. To summarize, the background data appear to follow both a normal and lognormal distribution, but Site A data do not appear to follow either a normal or lognormal distribution. A dilemma exists regarding the distribution of the background data—is it normal or lognormal? As the log transformation did not appreciably improve the normality of the data set, it would be advisable not to perform the transformation.

[*] Appendix F.

Table 3-4.
Results of the Shapiro-Wilk Test of Normality and Lognormality for Chromium Surface Soil at Site A and Background

Area	Testing for Normality or Lognormality?	Number of Results	*CV*	Shapiro-Wilk Test Statistic, *W*	Critical Values (from Table B-20 of Appendix B)			*p* value for Shapiro-Wilk Test (from statistical software)	Conclusion: Is there evidence that the data are Normally or Lognormally Distributed? Yes/No
					Critical Value based on 0.05 level of significance, $W_{0.05}$	Critical Value based on 0.10 level of significance, $W_{0.10}$	Critical Value based on 0.50 level of significance, $W_{0.50}$		
Site A	Normality	24	0.5687	0.627	0.916	0.930	0.963	<0.0001	No
Site A	Lognormality	24	0.2426	0.791	0.916	0.930	0.963	0.0002	No
Background	Normality	16	0.1093	0.963	0.887	0.906	0.952	0.7177	Yes
Background	Lognormality	16	0.07041	0.958	0.887	0.906	0.952	0.6308	Yes

Table 3-5.
Results of the Studentized Range Test of Normality and Lognormality for Chromium Surface Soil at Site A and Background

Area	Number of Results	Test of Normality (based on original data)			Test of Lognormality (based on log-transformed data)		
		Ratio of Range of Results and Standard Deviation	Critical Values from Table B-21 of Appendix B, assuming a 0.05 level of significance	Conclusion: Is there evidence that the data are Normally Distributed? Yes/No	Ratio of Range of Results and Standard Deviation	Critical Values from Table B-21 of Appendix B, assuming a 0.05 level of significance	Conclusion: Is there evidence that the data are Lognormally Distributed? Yes/No
Site A	24	4.586	(3.308, 4.666)*	Yes	4.400	(3.308, 4.666)*	Yes
Background	16	3.278	(3.01, 4.24)	Yes	3.317	(3.01, 4.24)	Yes

*Critical Values for n = 24 are based linear interpolation of critical values from n = 20 and n = 25.

3.13.15. If a background value, such as a UTL, and other summary statistics are desired to characterize the background data set, then the assumed distribution should fit the data as much as possible. With respect to this objective, it would be more appropriate to define background as following a normal distribution because the Shapiro-Wilk test shows more evidence of normality than lognormality. Comparing the Shapiro-Wilk test's critical value or associated *p* value from the original data and from the log-transformed data is a reasonable approach for discerning which distribution is more appropriate and has more evidence of following a normal or log-normal distribution.

3.13.16. The first objective for this case study is to determine whether chromium contamination at Site A, on the average, exceeds a regulatory threshold value. As it cannot be assumed that the Site A data set is either normal or lognormal, a nonparametric test (e.g., the Wilcoxon signed rank test for the median as discussed in Appendices H and M) must be used to compare the Site A data to the regulatory threshold.

3.13.17. The second objective is to determine whether chromium exceeds background. Though the background data set could be reasonably assumed to be either normal or lognormal, this assumption could not be made for the Site A data set. As the Site A data set is neither normal nor lognormal, a *parametric two-sample test*[*] cannot be used to compare the Site A data set to the background data set (for example, to determine if the mean concentration at Site A exceeds the mean background concentration). Both data sets must follow the same distribution to use a parametric test. For example, both the background and site data sets must both be normally or lognormally distributed. As data from Site A does not follow a normal or lognormal distribution, only nonparametric tests such as the *Wilcoxon rank-sum test** can be used to compare the Site A and background data sets.

3.13.18. This case study illustrates the value of background data in project decision-making. The application of background data in identifying contaminants for inclusion in the risk assessment is presented in the following section. The data in the preceding discussion may be used as sample data to apply some of the nonparametric tests in Appendix M.

Section III
Risk Assessment

3.14. Introduction. Perhaps more than any other area in the CERCLA project life cycle, assessing site risk relies on statistics. Many of the techniques described in several of the appendices apply in quantifying and assessing risk at a hazardous waste site. The components of a risk assessment discussed in this report are:

[*] Appendices M and N.

- Identifying contaminants of potential concern (COPCs).
- Calculating exposure point concentrations (EPCs).

Statistics enter into risk assessment in one additional major area—the calculation of exposure levels. Specifically, a baseline human health risk assessment requires estimation of a reasonable maximum exposure (RME), and a central tendency exposure (CTE). The former relies on 95% *upper confidence level (UCL)* values for exposure parameters, and the latter on the mean of the exposure parameters. In either case, the exposure parameters are generally provided by EPA guidance, such as the *Exposure Factors Handbook* (USEPA, 1997). For all practical purposes, the environmental scientist will not need to statistically evaluate these parameters and, consequently, their derivation is not discussed here. However, understanding the concepts presented in Appendix E is very useful in deconstructing the data evaluations presented in the *Exposure Factors Handbook* (USEPA, 1997).

3.14.1. *Identification of Contaminants of Potential Concern for Risk Assessment.* Not all chemicals detected at a site are typically included in the quantification of risk. Those chemicals retained in the risk assessment are the COPCs. Note that the COPCs are media-specific; COPCs are evaluated for air, surface soil, subsurface soil, groundwater, sediment, surface water, and any other medium sampled in the RI at each site.

3.14.1.1. Chemicals are typically screened against background or other criteria (established by ARARs) and a subset is selected for inclusion in the risk calculations. Some of the screening criteria, other than background levels, include drinking water MCLs, or secondary MCLs, RBCs, and *Toxic Substance Control Act* (TSCA) values for PCBs (polychlorinated biphenyls) in soil. In addition, inorganics that are essential human nutrients (e.g., iron, potassium, magnesium, sodium, and calcium) may be excluded from the quantitative risk analysis in most cases. (ARARs are identified in the planning stage of the RI.)

3.14.1.2. Both qualitative and quantitative statistical evaluations are frequently performed to identify COPCs. A qualitative evaluation is initially conducted to determine whether select potential analytes of concern can be eliminated from future investigation; a statistical evaluation is subsequently done for a more in-depth look at of contaminants that were not eliminated during the qualitative assessment.

3.14.1.3. For example, for the qualitative evaluation of the data, if a chemical is detected infrequently in the sample data set, and is not considered to be associated with historical waste handling at a site, it may be screened out as a COPC. However, it is essential to use site-specific information before discarding such a chemical, as infrequently detected compounds may also represent hot-spots, depending on the sampling strategy used at the site. For every chemical detected at least once, the maximum detected concentration is compared to the chemical- and

medium-specific screening criterion. Chemicals with higher concentrations than their criteria are generally retained for quantitative evaluation in the risk assessment.

3.14.1.4. Contaminants that lack ARARs (usually because toxicity information does not exist) are retained as COPCs in the risk assessment and discussed in the uncertainty section of the report. One-sample tests for contaminants where the maximum exceeds the risk-based screening limit may be used to determine whether the mean is statistically less than the screening limit, even though a single value exceeds the screening limit. Anthropogenically derived contaminants (such as PAHs) that occur at concentrations below background levels are still retained in the risk assessment if they exceed ARARs. If the risk assessment indicates that such contaminants are a primary contributor to total risk at a site, then a quantitative statistical comparison with background (e.g., using appropriate two-sample statistical tests) would be done and the results would subsequently be discussed in the risk characterization at the end of the assessment.

3.14.2. *Calculating Exposure Point Concentrations*. For risk assessment, means and standard deviations are typically calculated as the basis for EPCs and as the basis for deriving UTLs for the background comparisons. However, the mean and standard deviation will frequently be inappropriate measures of central tendency and dispersion when the data are not normally distributed or a large portion of the data consists of non-detects. Under these circumstances, means and standard deviations should not be used to perform statistical evaluations. Before statistically valid means and standard deviations can be calculated, tests for normality should be conducted and non-detects must be appropriately addressed.

3.14.2.1. The EPC is used to calculate a COPC's carcinogenic risk and non-carcinogenic hazard index. It represents the concentration a receptor is likely to encounter. The USEPA requires the EPC to be a conservative estimator of central tendency—the 95% upper confidence limit (UCL) of the sample arithmetic mean concentration (OSWER 92-856-03, EPA 68-W0-0025). The 95% UCL is the concentration that, when calculated repeatedly for randomly drawn samples, equals or exceeds the true mean 95% of the time.

3.14.2.2. Calculating rigorous, statistically valid 95% UCLs requires that data be distribution tested and that non-detects be treated properly. Procedures for this are provided in Appendix H. Some of the older (pre-2000) RCRA and CERCLA guidance for calculating the UCL are outdated (and hence, are not recommended); modifications and updates are provided with the goal of improving scientific defensibility. Appendix G presents the most recent acceptable methods for estimating the 95% UCL at 95% confidence.

3.14.2.3. Calculating EPCs at a CERCLA site brings together many of the statistical procedures described in the attached Appendices. The correct steps are, in general, as follows

3.14.2.3.1. Identify the nature of the censoring limit and the proportion of censored values and substitute proxy values as directed in Appendix R.

3.14.2.3.2. Identify outliers as discussed in Appendices I and M.

3.14.2.3.3. Perform distribution testing as detailed in Appendix F.

3.14.2.3.4. Depending on the outcome of these steps, calculate the 95% UCL as directed in Appendix G.

3.14.2.4. Unfortunately, there are many pitfalls along the way, and this process does not always lead to a simple result. In part, this is attributable to the use of or adherence to older USEPA guidance. In particular, USEPA guidance for substituting for censored data is addressed in many separate risk assessment documents. In earlier documents, substituting one-half the detection limit is supported. Appendix E provides insight on the deficiency in this approach. In addition, even if the risk assessor has performed all of the statistical procedures, USEPA guidance for EPCs states that if a 95% UCL exceeds the maximum value of a compound detected at a site, the maximum should be substituted. This has the dissatisfying attribute of being completely *ad hoc*, giving rise to unquantifiable and unacceptable uncertainties for risk assessment decisions.

3.14.3. *Uncertainty Quantification*. A required element in a baseline human health risk assessment is to evaluate uncertainty for decisions. Statistical techniques alone will be unable to account for all sources of uncertainty in a risk assessment and a qualitative approach is normally taken. For example, there will be uncertainty in the risk assessment for analytes for which toxicity data do not exist, and the quantification of such uncertainty is not possible.

3.14.3.1. In risk assessment, uncertainty stems primarily from the following three sources.

3.14.3.1.1. Errors in the estimate of contaminant concentration.

3.14.3.1.2. Errors in the estimate of toxicity.

3.14.3.1.3. Errors introduced by large numbers of assumed values in the risk assessment formulations, which are by definition and intent very conservative.

3.14.3.2. In practical terms, there is little that can be done about the uncertainty in estimates of toxicity. The studies upon which toxicity data are based are taken “as is” simply because of the scarcity of available studies. Uncertainty in the assumptions employed in the risk assessment can sometimes be addressed, but only to a limited extent. An example for how the uncertainties listed in subparagraph 3.14.3.1.3 were taken into account is presented in Case Study 6.

3.14.3.3. Most statistical evaluations implicitly assume the absence of bias. The uncertainty predominantly depends on the distribution of field measurements. Even in the case of risk screening, as demonstrated in Chapter 2, we have seen that it is possible to qualitatively assess

the uncertainty of individual sample/analytical results before comparing those results to fixed threshold values using analytical QC information. For example, QC data can potentially be used to identify the direction of bias and to estimate the magnitude of the bias associated with a set of analytical results. This is illustrated in Case Study 6. It is also possible to make similar estimates of variability which may affect decision-making, as illustrated in Case Study 7.

3.14.3.4. The error introduced into the risk assessment by the uncertainty associated with each of the various assumptions and reference values is more likely multiplicative rather than additive, such that the calculated risk is conservative to an extraordinary degree. Consider, for instance, some components of a soil dermal absorption scenario. The risk assessor calculates an EPC, which represents the 95% UCL of the mean. Then, the skin area exposed to the contaminant is based on an upper 95% confidence level of all the U.S. adult population from EPA OSWER 92-856-03. These are combined with, say, the default average exposure duration and frequency values which, again, are upper estimates from some population. Combining all of these upper estimates results in a risk evaluation that has a far higher confidence than 95%. The Risk Assessor and Project Manager are encouraged to identify every opportunity to use site-specific values in place of assumptions in risk assessment to reduce uncertainty in the results and, thus, more appropriately apply the limited remediation resources available.

3.14.3.5. One method for estimating the true mean and distribution of risk estimates is to use the recommended RME and CTE values of exposure parameters. This methodology is recommended in *Risk Assessment Guidance for Superfund* (RAGS). The result of looking at each input parameter using the CTE is to provide an estimate of risk near the mean of the estimated exposure scenario. The RME is considered to represent an upper estimate of site risk. An alternative method of quantifying the range in risk estimates is to use Monte Carlo simulations.

3.15. Case Study 6—Refining Risk Assessment Assumptions.

3.15.1. A risk assessment was to be done as part of a RCRA Facility Investigation (RFI) at a steel mill in Pennsylvania. The project team approached the EPA Remedial Project Manager (RPM) regarding using site-specific assumptions for some of the exposure factors in the risk assessment calculations. This was possible because the facility maintained excellent records of employee longevity, promotion, and work assignments. For this case study, the focus is on site-specific estimates of exposure duration, which enters into quantification of risk.

3.15.2. Under the assumptions given by the EPA for the worker exposure scenario in OSWER 92-856-03, the risk assessor is to assume that a given worker will be exposed for a period of 25 years. However, by reference to detailed employee records for the facility, the project team was able to demonstrate concretely on a facility-specific, job-specific, and location-specific basis, the actual average lifetime exposure duration for the various site areas under study. Employing these actual values, which were approximately 3 to 5 years rather than 25 years, greatly reduced the exposure duration. More importantly, the site-specific value reduced the uncertainty

in the calculated lifetime risk. Using this lower value allowed the steel mill owner to limit the number of site areas proceeding to the Corrective Measures Study phase of the project.

3.16. Case Study 7—Direction and Magnitude of Bias. As part of a property transfer in Baltimore, Maryland, the project team was asked to estimate reserves that the seller would have to put in escrow against the potential need for site clean-up, before the seller would accept transfer of the property. For this case study, petroleum hydrocarbon contamination will be discussed.

3.16.1. The project team decided to divide the relatively small site into four quadrants and collect one composite sample from each to assess the potential need for remediation in each quadrant. The analytical results obtained from the laboratory were as follows:

Quadrant 1	1200 mg/kg	Quadrant 3	756 mg/kg
Quadrant 2	101 mg/kg	Quadrant 4	138 mg/kg

3.16.2. With the state's action level set at 100 mg/kg, it appeared that the seller would be required to reserve funds against a potential soil removal for the entire site. However, a review of the quality control data associated with the analytical results displayed significant potential bias.

3.16.3. A normal calibration curve was developed for the gas chromatograph used in the analysis that met method criteria for linearity. The laboratory then analyzed an Initial Calibration Verification (ICV) using a standard from an alternative source from that employed in the calibration. The ICV was essentially a blank spike set at the midpoint of the calibration curve. The result of this analysis was a percent recovery (%R) of 168%, which was within the acceptance limits provided with the standard by the manufacturer.

3.16.4. However, in its simplest form this QC result indicates that if the laboratory introduced the equivalent of 100 mg/kg of total petroleum hydrocarbons (TPH) into the analytical system, they would get a reported result of 168 mg/kg. This observation, applied to the results reported for the site, removed two of the four quadrants from further consideration, reducing the required reserves by half.

Section IV
Probabilistic Risk Assessments Monte Carlo Simulations

3.17. Introduction. The implementation of probabilistic risk assessment for environmental projects is beyond the scope of this document; however, a brief overview of the procedures is presented here. Monte Carlo simulation, the most common technique used for probabilistic assessments, is a statistical technique in which outcomes are produced using randomly selected values for input variables that possess a range of possible values. In some cases, a known probability distribution can be assigned to each input variable. By repeating the calculation many, many times, Monte Carlo simulations create a population of results representing (in theory) the

full range of possible outcomes and the likelihood of each. For example, when Monte Carlo simulation is used in risk assessment, risk is expressed as a distribution of possible values rather than a single point value.

3.17.1. There are two major practical limitations to the application of Monte Carlo simulations in general: i) it can be costly, and ii) few people are sufficiently qualified to do it. The EPA has also written a guidance document for probabilistic risk assessment titled *RAGS Volume 3 Part A: Process for Conducting Probabilistic Risk Assessment* (EPA 540-R-02-002) available at http://www.epa.gov/oswer/riskassessment/rags3a/index.htm. An EPA Region 3 publication (EPA 903-F-94-001) identified several technical limitations that preclude the Agency from relying on Monte Carlo simulations (http://www.epa.gov/reg3hwmd/risk/human/info/guide1.htm).

3.17.1.1. Software is unable to distinguish between measurement variability and lack of knowledge. Some input parameters are for well-described differences among individuals—these differences are variability. Other factors, such as frequency and duration of trespassing, are simply unknown, and assuming a distribution for them is ad hoc. But the simulated distribution of unknowns is presented in computer output as variability. The accuracy of the distributional assumptions limits the accuracy of the simulation.

3.17.1.2. Software is unable to account for sample dependency (e.g., spatial and temporal correlations for sample locations). However, this limitation also applies to all classical statistical methods (e.g., the methods predominantly discussed in this document and in EPA environmental statistical documents such as the QA-G4 and GA-G9 guidance documents). In classical statistics, the assumption of independence highly influences the applicability of a technique—the same limitation applies here.

3.17.2. In most statistical evaluations (excluding geostatistics), environmental scientists are resigned to the limitations of classical statistics for environmental data. The same is true for Monte Carlo simulations. Though Monte Carlo simulations require sample independence, the approach can be advantageous. The primary advantage is that it accounts for a range of input values and outputs a range of outcomes (such as risk values) with associated probabilities. Although a Monte Carlo approach is currently not recommended or required by the EPA, the approach may be beneficial for some projects. There are applications of such simulations. Moreover, future scientists may learn how to overcome some of the limitations and eventually develop reasonable and inexpensive computer applications.

3.17.3. Applications of Monte Carlo simulation are more prevalent in groundwater modeling than any other current environmental application. Case Study 8 shows how a Monte Carlo simulation of groundwater contamination was used to perfect a remedy.

3.18. Case Study 8—Monte Carlo Simulation in Remedial Alternative Selection

3.18.1. Monte Carlo analysis was coupled with decision tree analysis for a study site in Nebraska where the groundwater was contaminated with trinitrotoluene (TNT). The extent of TNT contamination was characterized during an RI. Three pump-and-treat alternative remedial actions were developed for the FS. The maximum concentration of TNT remaining in the saturated zone at the end of each alternative project lifetime was determined stochastically using a Monte Carlo model. The Monte Carlo model randomly generated values for site information for initial mass concentration, hydraulic conductivity, and retardation coefficient. Then these randomly generated fields were sampled and the output was combined into sets or ensembles. Probability functions were fitted to the output ensembles with the maximum simulated TNT concentrations. Because each of the treatment alternatives was associated with a different set of possible maximum concentrations, the Monte Carlo simulation made it possible to identify the optimal alternative quantitatively by analyzing the output ensembles for each alternative.

3.18.2. Applying Monte Carlo simulations requires the technical support of a specialist in this area; detailed methodologies are beyond the scope of this Manual. The technique does rely on the power of randomly generated data sets and the optimization of conditions based on the simulation.