

REPRESENTATIVE CONCENTRATION WHITE PAPER

1. OVERVIEW

Human health and ecological risk assessments assume that receptors are exposed to chemicals of potential concern (COPCs) over a defined spatial and temporal scale and through various exposure pathways. In the human health risk assessment, an exposure point concentration is the COPC concentration in an exposure medium at the time and location where a receptor may contact that medium. Exposure point concentrations are calculated based on representative concentrations. The distinction between exposure point concentrations and representative concentrations for the human health risk assessment is that representative concentrations are “representative” of the exposure to a particular media, whereas exposure point concentrations also encompass modeled concentrations in other exposure media. Representative concentrations are used directly in the ecological risk exposure assessment. Ideally, representative concentrations are estimated from sampled media spatial scales consistent with the receptors' exposure scales.

The risk assessments include an evaluation of the reasonable maximum exposure (RME) and the central tendency exposure (CTE). An RME scenario assesses risk to individuals whose behavioral characteristics may result in much higher potential exposure than seen in the average individual. A CTE scenario assesses potential risk to an average member of the population. The inclusion of both RME and CTE calculations provides a semi-quantitative measure of the range of expected risks that may occur under a particular exposure scenario. The CTE and RME are a simple form of probabilistic risk assessment and provides risk managers with an estimate of the mean and upper percentile of estimates of exposure. RME scenarios should not identify every input parameter as an upper-bound estimate, otherwise conservatism will compound into an unrealistic upper-bound risk estimate.

Generally, representative concentrations are needed that characterize both the RME and CTE scenarios. In Draft A of the River Corridor Baseline Risk Assessment (RCBRA), representative concentrations were calculated using means as CTE estimates and 95% upper confidence limits (UCLs) on the mean as RME estimates. This approach was applied across all data sets and analytes, although in several cases data had not been collected with the intention of supporting UCL calculations. In addition, the UCLs were based on a default assumption of lognormal and were calculated using Land's method. As a consequence, some UCL values in the Draft A RCBRA were unrealistically large (see EPA 2007 for more information on Unreliable UCL, Unstable UCL, Unrealistic UCL).

The statistical methods for calculating the UCL values for use in risk assessment have evolved over time. Originally, U.S. Environmental Protection Agency (EPA) guidance (EPA 1992) provided methods for calculating UCLs based on data derived from a normal and lognormal statistical distribution; this guidance also allowed for using the maximum concentration if the UCL was larger than maximum value. This original approach was deficient in several ways and could reward responsible parties for collecting small samples of variable concentration data. Starting in 2002, EPA has released versions of the ProUCL software to address technical and computational issues associated with calculating UCLs from environmental data. Many of these

techniques have been practical with the advances in computer hardware and software since the original EPA guidance was released in 1992.

A revised process for calculating analyte concentrations in environmental media is needed so that RME risks can be calculated. In general, the process described in the following sections follows EPA guidance as provided in the ProUCL Version 4 User Guide and Technical Guide (EPA 2007). Several issues that need to be considered for determining the most appropriate methods for estimating representative concentrations for CTE and RME scenarios:

- What was the intended use of the sample results (what were the data quality objectives)?
- How many sample results are available for the exposure unit?
- Are the data censored (are there nondetect sample results)?
- What estimation methods are mathematically stable for the data being evaluated and therefore provide reasonable estimates of the mean and upper bound on the mean?

The *Washington Administrative Code* (WAC) also provides methods for calculating representative concentrations. The WAC methods can be found in the compliance monitoring sections of the code (e.g., WAC 173-340-740(7)); these methods use the lognormal distribution assumptions and other statistical methodologies no longer advocated in EPA guidance. However, the WAC does permit use of “other statistical methods approved by the department.” Before exploring these issues, the data sets and exposure unit areas for the ecological and human health risk assessments are defined in Section 2.

2. DATA SETS AND EXPOSURE UNITS

Every representative concentration used in the risk assessment has an associated spatial scale to which the underlying data apply. Risk assessment results derived from the representative concentration pertain to that spatial scale. Spatial scales are commonly designated as exposure units, or exposure areas, in a risk assessment. These are commonly defined on a volumetric basis for abiotic media. For example, RCBRA multi-increment soil samples represent the top 15.2 cm (6in.) of soil within an area of approximately 1 ha. Sediment samples are collected from the biologically active 0 – 10.2 cm (0 to 4 in.) zone and, in the case of samples collected from operational areas, represent conditions within the area where groundwater contaminant plumes enter the Columbia River. For biotic media, an exposure unit is a function of the home range of the sampled species. This can vary from less than 1 ha (in the case of certain waste sites) to several hectares or more, such as for avian species like kingbirds. A crosswalk of spatial scale and the data sets used to evaluate representative concentrations for the human health and ecological risk assessments is provided in Table 1. In the case of the RCBRA data, the ultimate spatial scale for calculation of representative concentrations depends upon the outcome of data analysis indicating the presence of significantly different concentrations among sampling locations. As detailed in Section 4, the spatial scale influences the method use to estimate representative concentrations largely through the number of samples, the number of detected samples, and the skewness of the sample dataset segmented at a particular scale.

Table 1. Summary of Representative Concentration Data Sets.

Data Group ^a	Sample Medium	Human Health Exposure Medium	Ecological Exposure Medium	Local Area Statistic and Spatial Scales	Broad Area Statistic and Spatial Scales
CVP/RSVP	Shallow zone soil	X	Previously evaluated in SLERA	CTE/RME; Individual waste site; 0 – 4.6 m	Not applicable
	Deep zone soil	X	Previously evaluated in SLERA	CTE/RME; Individual waste site; >4.6 m	Not applicable
Groundwater	Shallow aquifer water	X	Not applicable	TBD, well	TBD, well grouping
RCBRA Upland	Soil	X	X	CTE/RME; Approx 1 ha	Varies by operating area
	Plants	X	X	CTE; 2 samples	Varies by operating area
	Invertebrates	Not used	X	CTE; 1 sample	Varies by operating area
	Small mammals	Not used	X	CTE; 1 sample	Varies by operating area
RCBRA Riparian	Soil	X	X	CTE/RME; Approx 0.1 ha	Varies by operating area
	Plants	Not used	X	CTE; 2 samples; approx 0.1 ha	Varies by operating area
	Invertebrates	Not used	X	CTE; 1 sample; approx 0.1 ha	Varies by operating area
	Small mammals	Not used	X	CTE; 1 sample; approx 0.1 ha	Varies by operating area
	Birds	Not used	X	CTE; Area varies	Varies by operating area
RCBRA Near-shore	Sediment	Not used	X	CTE; 1 sample	Varies by operating area
	Surface water	X	X	CTE; 1 sample	Varies by operating area
	Pore water	X	X	CTE; 1 sample	Varies by operating area
	Macroinvertebrates	X	X	CTE; 1 sample	Varies by operating area
	Clams	X	X	CTE; 1 sample	Varies by operating area
	Sculpin	Not used	X	CTE; 1 sample	Varies by operating area
Other (environmental surveillance, etc.)	Soil	X	X	Not applicable	Varies by operating area
	Plants	X	X	Not applicable	Varies by operating area
	Elk and other game	X	Not applicable	Apply operating area data	Varies by operating area
	Salmon and other food fish	X	Not applicable	Apply operating area data	Varies by operating area

^a For all data groups except for “Groundwater” and “Other”, representative concentrations are also calculated using analogous Reference Area samples.

CTE = Calculated Toxicity Equivalence

CVP = Cleanup Verification Package

RCBRA = River Corridor Baseline Risk Assessment

RME = Reasonable Maximum Exposure

RSVP = Remaining Sites Verification Package

SLERA = Screening Level Ecological Risk Assessment

TBD = To Be Determined

For some of the combinations of data group and sample medium shown in Table 1, sampling was not always performed with the intention of calculating a UCL. In these cases, there may be too few samples available to estimate the RME. Sometimes, there are few samples because they were collected to represent “worst-case” conditions (such as cleanup verification package/remaining site verification package [CVP/RSVP] samples biased based on field screening). Other times, the data are used within an assessment context that considers multiple lines of evidence, and protectiveness is ensured as a function of the overall process.

3. TREATMENT OF CENSORED VALUES

A nondetect is an analytical sample result where the concentration is deemed to be lower than could be reliably identified and quantified using the method employed by the analytical laboratory. A value is reported and a qualifier assigned indicating that the sample concentration was smaller than that value. The data are essentially censored at this value. Thus, nondetect results are referred to as censored data. Nondetects may correspond to concentrations that are actually or virtually zero, or they may correspond to values that are larger than zero but which are below the laboratory’s ability to provide a reliable measurement. All approaches to working with nondetects use substitution values to estimate the sample results that might have been obtained with a more sensitive analytical method. Some of the most widely used methods are substitutions of the nondetects with the detection limit, half the detection limit (DL/2), or zero. EPA’s recently released ProUCL Version 4 Technical Guide (EPA 2007) “strongly recommends avoiding the use of the DL/2 method even when the percentage (%) of non-detects is as low as 5%-10%.” Two approaches for handling nondetects discussed in the ProUCL guidance include Kaplan-Meier estimation and Regression on Order Statistics (ROS).

Kaplan-Meier is a nonparametric approach for estimating the mean and standard deviation of censored data that is commonly used in survival analysis. The Kaplan-Meier method provides an estimate of the sample distribution functions adjusting for censored data. The Kaplan-Meier estimate of the sample distribution function is calculated as follows (EPA 2007):

$$\begin{aligned}\tilde{F}(x) &= 1, & x \leq x'_n \\ \tilde{F}(x) &= \prod_{j=x'_j > x} \frac{n_j - m_j}{n_j} & x'_1 \leq x \leq x'_n \\ \tilde{F}(x) &= \tilde{F}(x'_1) & x_{\min} \leq x \leq x'_1 \\ \tilde{F}(x) &= 0 & 0 \leq x \leq x_{\min}\end{aligned}$$

where: x is a vector of n samples

x' is a vector of n' distinct **detected** samples

m_j is the number of **detects** at x'_j

n_j is the number of samples $\leq x'_j$

x_{\min} is the minimum value of x .

The Kaplan-Meier estimate of the population mean is given by

$$\hat{\mu} = \sum_{i=1}^{n'} x'_i [\tilde{F}(x'_i) - \tilde{F}(x'_{i-1})] \text{ with } x'_0 = 0.$$

The Kaplan-Meier estimate of the standard error of the mean is given by

$$\hat{\sigma}_{SE}^2 = \frac{n-k}{n-k-1} \sum_{i=1}^{n'-1} a_i^2 \frac{m_{i+1}}{n_{i+1}(n_{i+1} - m_{i+1})}$$

where k = the number of nondetects and

$$a_i = \sum_{j=1}^i (x'_{j+1} - x'_j) \tilde{F}(x'_j), \quad i = 1, 2, \dots, n' - 1.$$

The Kaplan-Meier parameter estimates can then be used to estimate the UCL parametrically or the mean estimator can be used in a bootstrap resampling algorithm.

Alternatively, ROS is a parametric approach for imputing nondetect concentrations. ROS estimates a linear model of the detected sample values versus the quantiles from a hypothesized probability distribution and then uses this linear model to assign values for the nondetects. The quantiles can be based on an assumed distribution such as a normal, lognormal, or gamma. The first step in computing the quantiles is to compute the plotting positions or percentiles. For example, in the case of one detection limit for the gamma distribution, ProUCL computes the plotting position for the i^{th} ordered sample value as $(i-1/2)/n$. The gamma quantiles are then computed using the probability statement $P(X \leq q(i)) = (i-1/2)/n, i = 1, 2, \dots, n$, where X represents a gamma random variable. Details for computing plotting positions for multiple detection limits can be found in Helsel (2005). Once the model is fit, values for the nondetects can be imputed concentrations and combined with the detected concentrations and a mean and UCL can be estimated using a parametric UCL method.

Given the use of linear model at least two detected observations are needed to estimate the model. However, the reliability of the assignment is highly dependent on the number of detected values. ProUCL recommends at least 8 to 10 detected values are needed for a reliable model.

Both Kaplan-Meier and ROS can handle multiple detection limits. With multiple detection limits, ROS can assign values for nondetects that are higher than some detected values, which is not the case with Kaplan-Meier. Although this may appear to be a disadvantage, it can also be seen as an advantage given that there is a nonzero probability that a nondetect value is actually above the detection limit. Other trade-offs between Kaplan-Meier and ROS are typical of the general trade-offs between parametric and nonparametric approaches. When parametric assumptions are met, parametric methods are considered to be more powerful while if the assumptions are not met then nonparametric methods are considered more powerful. As detailed in Section 4, both the Kaplan-Meier and ROS approaches for handling censored values will be employed in estimating means and confidence and levels and, subsequently, the values used as the CTE and RME.

4. CALCULATING METHODOLOGY FOR MEANS AND UPPER CONFIDENCE LIMITS ON THE MEAN

This section provides the technical approach for calculating means and UCLs on the mean for RCBRA. The decision logic for choosing an appropriate statistical method is largely based on the number of detected samples and the skewness of the available samples for the spatial scale of

interest (e.g., site versus operational area). The process is largely based on EPA guidance as presented in ProUCL (EPA 2007). In practice, deviations from ProUCL are unavoidable when making calculations for a large number of COPCs at various scales. For example, in some cases ProUCL recommends bootstrapping the Kaplan-Meier estimate of the mean. However, the Kaplan-Meier estimate of the mean is undefined when a bootstrap sample draws all nondetects.

In *Calculating Upper Confidence Limits for Exposure Point Concentrations at Hazardous Waste Sites* (EPA 2002), EPA recommends using the average concentration to represent “a reasonable estimate of the concentration likely to be contacted over time” (EPA 1989) and “because of the uncertainty associated with estimating the true average concentration at a site” recommends that the 95% UCL on the mean be used for assessing a reasonable maximum exposure. There are many parametric and nonparametric methodologies available for calculating UCLs. Representative concentration estimation for RCBRA will calculate the mean and UCL using both parametric and nonparametric methodologies in cases in which the number of detected samples, n' , is greater than 4. To provide robustness to distributional assumptions, outliers, and methodology instabilities, the median of the calculated means and the median of the calculated UCLs will be used as the CTE and RME, respectively.

4.1 Approach for Small Number of Detected Values

The following is a summary of the approaches for calculating representative concentrations for detect samples sizes, n' , from 1 to 4:

- $n' = 1$; then result is the CTE, no RME
- $n' = 2$; then maximum detect is the CTE, no RME
- $n' = 3$ or 4; then the average is the CTE, maximum detect is used as RME. This will be shown on the EPC listing and will be clearly explained in the text.

4.2 Approach for Large Number of Detected Values

4.2.1 Parametric Approach

For parametric approaches, the particular mean and UCL estimation methodology used will depend on the distributional assumptions made for the population that generated the data. ProUCL recommends avoiding “the use of a lognormal model even when the data appear to be lognormally distributed” because “its use often results in incorrect and unrealistic statistics of no practical merit” (EPA 2007). ProUCL further notes that “the gamma distribution is better suited to model positively skewed environmental data sets” based on the following:

- The conclusion that use of a gamma distribution results in reliable and stable UCL values for datasets without nondetects
- The fact that there is no need to transform the data and back-transform the resulting statistics.

The gamma distribution also has the advantage that it can also be used to model normally distributed data. Given the above, the assumption will be made that concentration data are gamma distributed and the gamma shape and scale parameters can be used to estimate the UCL based on the following methods (EPA 2007):

- Student t 95% UCL
- Approximate gamma 95% UCL
- Adjusted gamma 95% UCL.

Table 2 provides the parameter estimation decision logic based on the gamma shape parameter k , and the number of observations, n (EPA 2007). For situations that include nondetects, a gamma ROS approach will be used to impute concentrations prior to mean and UCL estimation.

Table 2. Parametric Computation of a UCL of Mean of a Gamma Distribution with Shape Parameter, k , and Sample Size, n .

k	Sample Size, n	Method
$k \geq 0.5$	For all n	Approximate gamma 95% UCL
$0.1 \leq k < 0.5$	For all n	Adjusted gamma 95% UCL
$k < 0.1$	$n < 15$	95% Hall's bootstrap UCL
	$n \geq 15$	Adjusted gamma 95% UCL

Source: EPA 2007, Table 2.
UCL = Upper Confidence Limit

4.2.2 NonParametric Approach

In combination with the previously discussed parametric approaches based on the gamma distribution, the mean and the UCL will also be estimated nonparametrically following the decision logic in Table 3. In the case of nondetects the mean (and standard deviation for the Chebyshev approaches) will be estimated using the Kaplan-Meier estimator.

Table 3. NonParametric Computation of 95% UCL of the Mean with Kaplan-Meier Estimation of the Mean and Standard Deviation in Cases with Non-Detects. (2 Pages)

σ	Sample Size, n	% Non detects	Method
$\sigma \leq 0.5$	For all n	All	95% UCL based on Student's t
$0.5 < \sigma \leq 1.0$	$n \leq 50$	0-20 %	95% Chebyshev
	$n > 50$		95% BCa Bootstrap
	All n	[20%, 40%)	95% BCa Bootstrap
	All n	$\geq 40\%$	95% UCL based on Student's t
$1.0 < \sigma \leq 1.5$	$n < 40$	$< 50\%$	99% Chebyshev UCL

Table 3. NonParametric Computation of 95% UCL of the Mean with Kaplan-Meier Estimation of the Mean and Standard Deviation in Cases with Non-Detects. (2 Pages)

σ	Sample Size, n	% Non detects	Method
	$n \geq 40$		97.5% Chebyshev UCL
	All n	>50%	95% BCa Bootstrap
$1.5 < \sigma \leq 2.0$	$n < 40$	<50%	99% Chebyshev UCL
	$n \geq 40$		97.5% Chebyshev UCL
	$n < 40$	$\geq 50\%$	97.5% Chebyshev UCL
	$n \geq 40$		95% Chebyshev UCL
$\sigma > 2.0$	$n < 50$	All	97.5% Chebyshev UCL
	$n \geq 50$		99% Chebyshev UCL

Source: EPA 2007, Table 3

UCL = Upper Confidence Limit

4.3 Summary

This above approach will provide both a parametric and nonparametric estimate of the mean and 95% UCL. The decision logics presented in Tables 2 and 3 are largely based on simulation studies while the strongest theoretical basis for calculating a UCL on the mean is the central limit theorem which points to a Student t estimation approach. ProUCL in some cases recommends estimating the UCL by two approaches and taking the maximum of the estimates. Here we propose a similar type of approach in which the median of the three UCLs estimated from Table 2, Table 3, and the Student t is used as the RME concentration used in the RCBRA Human Health Risk Assessment component. Additionally, cases in which any one of the three UCL estimates is different by some amount (e.g., 10% different) will be analyzed further. This should ensure that a reliable and reasonable estimate of the RME is produced.

5. CONCLUSIONS

Estimating representative concentrations in soil is addressed in WAC 173-340-740(7) (see Attachment A). WAC 173-340-740(7) requires the default assumption of lognormally distributed data and application of Land's method for calculating the UCL of the mean. Current EPA guidance, as presented in ProUCL (EPA 2007), does not recommend making the lognormal data assumption and recommends against using Land's method since it can produce unrealistic results. ProUCL guidance suggests that the gamma distribution is a flexible alternative distributional assumption to the lognormal for environmental data. The decision logic presented in Section 4, specifically in Table 2 and Table 3, is based on ProUCL guidance (EPA 2007).

WAC 173-340-740(7) also addresses methods for censored data. Censored data are frequently encountered in the RCBRA data and the methods currently recommended by EPA as presented

in ProUCL guidance (EPA 2007) are not consistent with the simple substitution methods described in WAC. The ROS and Kaplan-Meier approaches for handling censored data detailed in Section 3 are based on ProUCL guidance (EPA 2007).

In summary, the methods proposed for calculating representative concentrations do not follow the default procedures described in WAC, but being consistent with current EPA methodologies they credibly represent “other statistical methods approved by the department” as stated in the WAC 173-340-740(7).

6. REFERENCES

EPA, 1989, *Risk Assessment Guidance for Superfund (RAGS): Volume 1, Human Health Evaluation Manual (Part A)*, EPA/540/1-89/002, Interim Final, U.S. Environmental Protection Agency, Washington, D.C.

EPA, 1992, *Supplemental Guidance to RAGS: Calculating the Concentration Term*, Publication EPA 9285.7-081, U.S. Environmental Protection Agency, Washington, D.C.

EPA, 2002, *Calculating Upper Confidence Limits for Exposure Point Concentrations at Hazardous Waste Sites*, [OSWER 9285.6-10](#), U.S. Environmental Protection Agency, Washington, D.C.

EPA, 2007, *ProUCL Version 4.0 User Guide*, EPA/600/R-07/038, U.S. Environmental Protection Agency, Washington, D.C.

Helsel, D.R., 2005, *Nondetects and Data Analysis*, John Wiley and Sons, Inc, Hoboken, New Jersey.

Attachment A

WAC 173-340-740(7) Compliance Monitoring for Soil.

(a) Compliance with soil cleanup levels shall be based on total analyses of the soil fraction less than two millimeters in size. When it is reasonable to expect that larger soil particles could be reduced to two millimeters or less during current or future site use and this reduction could cause an increase in the concentrations of hazardous substances in the soil, soil cleanup levels shall also apply to these larger soil particles. Compliance with soil cleanup levels shall be based on dry weight concentrations. The department may approve the use of alternate procedures for stabilized soils.

(b) When soil levels have been established at a site, sampling of the soil shall be conducted to determine if compliance with the soil cleanup levels has been achieved. Sampling and analytical procedures shall be defined in a compliance monitoring plan prepared under WAC 173-340-410. The sample design shall provide data that are representative of the area where exposure to hazardous substances may occur.

(c) The data analysis and evaluation procedures used to evaluate compliance with soil cleanup levels shall be defined in a compliance monitoring plan prepared under WAC 173-340-410.

These procedures shall meet the following general requirements:

(i) Methods of data analysis shall be consistent with the sampling design. Separate methods may be specified for surface soils and deeper soils;

(ii) When cleanup levels are based on requirements specified in applicable state and federal laws, the procedures for evaluating compliance that are specified in those requirements shall be used to evaluate compliance with cleanup levels unless those procedures conflict with the intent of this section;

(iii) Where procedures for evaluating compliance are not specified in an applicable state and federal law, statistical methods shall be appropriate for the distribution of sampling data for each hazardous substance. If the distributions for hazardous substances differ, more than one statistical method may be required; and

(iv) The data analysis plan shall specify which parameters are to be used to determine compliance with soil cleanup levels.

(A) For cleanup levels based on short-term or acute toxic effects on human health or the environment, an upper percentile soil concentration shall be used to evaluate compliance with cleanup levels.

(B) For cleanup levels based on chronic or carcinogenic threats, the true mean soil concentration shall be used to evaluate compliance with cleanup levels.

(d) When data analysis procedures for evaluating compliance are not specified in an applicable state or federal law the following procedures shall be used:

(i) A confidence interval approach that meets the following requirements:

(A) The upper one sided ninety-five percent confidence limit on the true mean soil concentration shall be less than the soil cleanup level. For lognormally distributed data, the upper one-sided ninety-five percent confidence limit shall be calculated using Land's method; and

(B) Data shall be assumed to be lognormally distributed unless this assumption is rejected by a statistical test. If a lognormal distribution is inappropriate, data shall be assumed to be normally distributed unless this assumption is rejected by a

statistical test. The W test, D'Agostino's test, or, censored probability plots, as appropriate for the data, shall be the statistical methods used to determine whether the data are lognormally or normally distributed;

- (ii) For an evaluation conducted under (c)(iv)(A) of this subsection, a parametric test for percentiles based on tolerance intervals to test the proportion of soil samples having concentrations less than the soil cleanup level. When using this method, the true proportion of samples that do not exceed the soil cleanup level shall not be less than ninety percent. Statistical tests shall be performed with a Type I error level of 0.05;
- (iii) Direct comparison of soil sample concentrations with cleanup levels may be used to evaluate compliance with cleanup levels where selective sampling of soil can be reliably expected to find suspected soil contamination. There must be documented, reliable information that the soil samples have been taken from the appropriate locations. Persons using this method must demonstrate that the basis used for selecting the soil sample locations provides a high probability that any existing areas of soil contamination have been found; or
- (iv) Other statistical methods approved by the department.
- (e) All data analysis methods used, including those specified in state and federal law, must meet the following requirements:
 - (i) No single sample concentration shall be greater than two times the soil cleanup level. Higher exceedances to control false positive error rates at five percent may be approved by the department when the cleanup level is based on background concentrations; and
 - (ii) Less than ten percent of the sample concentrations shall exceed the soil cleanup level. Higher exceedances to control false positive error rates at five percent may be approved by the department when the cleanup level is based on background concentrations.
- (f) When using statistical methods to demonstrate compliance with soil cleanup levels, the following procedures shall be used for measurements below the practical quantitation limit:
 - (i) Measurements below the method detection limit shall be assigned a value equal to one-half the method detection limit when not more than fifteen percent of the measurements are below the practical quantitation limit.
 - (ii) Measurements above the method detection limit but below the practical quantitation limit shall be assigned a value equal to the method detection limit when not more than fifteen percent of the measurements are below the practical quantitation limit.
 - (iii) When between fifteen and fifty percent of the measurements are below the practical quantitation limit and the data are assumed to be lognormally or normally distributed, Cohen's method shall be used to calculate a corrected mean and standard deviation for use in calculating an upper confidence limit on the true mean soil concentration.
 - (iv) If more than fifty percent of the measurements are below the practical quantitation limit, the largest value in the data set shall be used in place of an upper confidence limit on the true mean soil concentration.
 - (v) The department may approve alternate statistical procedures for handling nondetected values or values below the practical quantitation limit.

(vi) If a hazardous substance or petroleum fraction has never been detected in any sample at a site and these substances are not suspected of being present at the site based on site history and other knowledge, that hazardous substance or petroleum fraction may be excluded from the statistical analysis.