Exposure Point Concentration Guidance for Discrete Sampling
Case Study 1: Pesticides in Soil

ATSDR received a petition from a citizen in a rural community regarding an elementary school recently constructed over a former crop-dusting airport active through the 1970s. The petitioner asked ATSDR to investigate the elementary school property for potential contamination. ATSDR accepted the petition and worked with state agencies to address the community's environmental health concerns.

The elementary school students use almost the entire school grounds as play areas, and members of the community use the school yard for various activities. A health assessor assigned the entire school campus as an exposure unit. During a recent environmental investigation, ten soil samples were collected across the school campus from 0 to 3 inches below ground surface (bgs) and tested for a suite of pesticides. At this hypothetical site, dieldrin, a persistent insecticide, was detected in all samples at levels shown in Table 1. ATSDR minimal risk levels (MRLs) for dieldrin are available for intermediate and chronic duration oral exposures and were chosen as comparison values by the health assessor.

Table 1
Environmental Sampling Data for Dieldrin[1]

Sample ID	Result (mg/kg)
Sample 1	0.11
Sample 2	1.5
Sample 3	4.5
Sample 4	5.0
Sample 5	0.66
Sample 6	0.83
Sample 7	3.5
Sample 8	1.6
Sample 9	1.5
Sample 10	3.1

Review the data provided above and answer the following questions.

---

[1] *NOTE: All sampling data provided are for a hypothetical site. Assume that the sampling data are of a known and high quality and are representative of actual exposure concentrations.*

Questions

1) What EPC statistic should the health assessor select for determining intermediate and chronic duration exposure at this site? Identify the criteria considered for making your selection.

Answer:

The environmental data have been screened for quality, the exposure unit is defined, and dieldrin was identified as a contaminant of concern. Therefore, the health assessor should follow the flow chart in Figure 1 from the EPC guidance document.
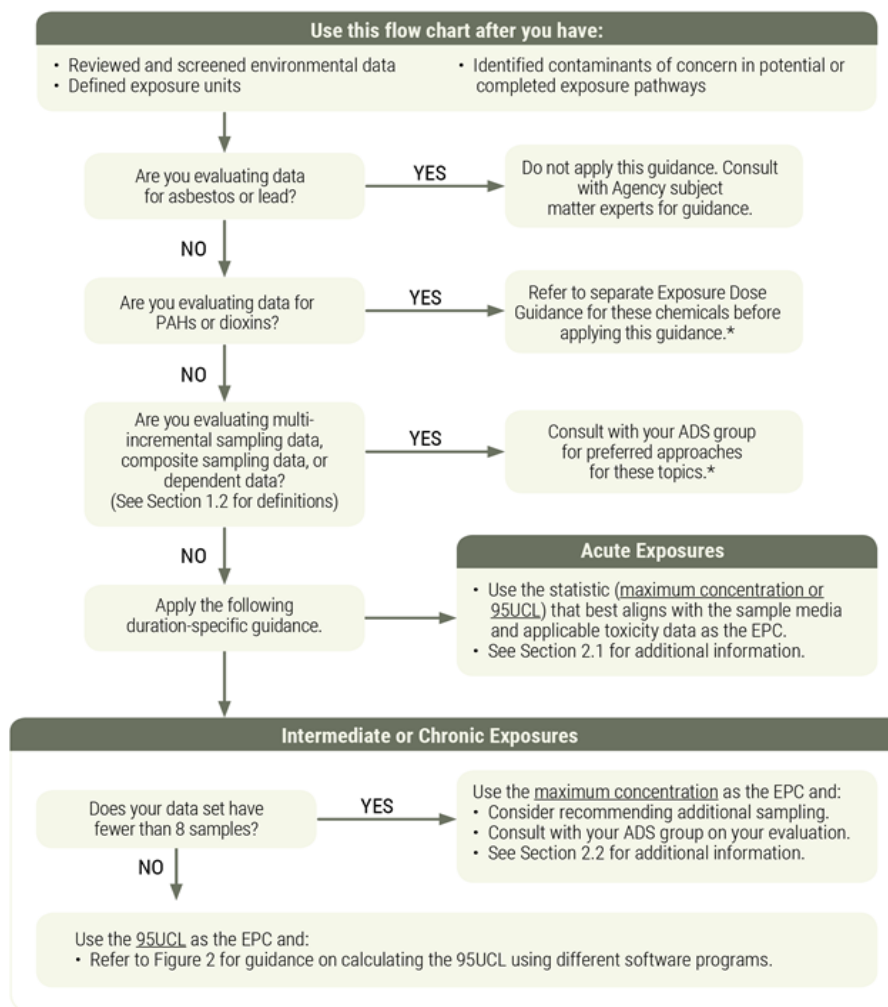
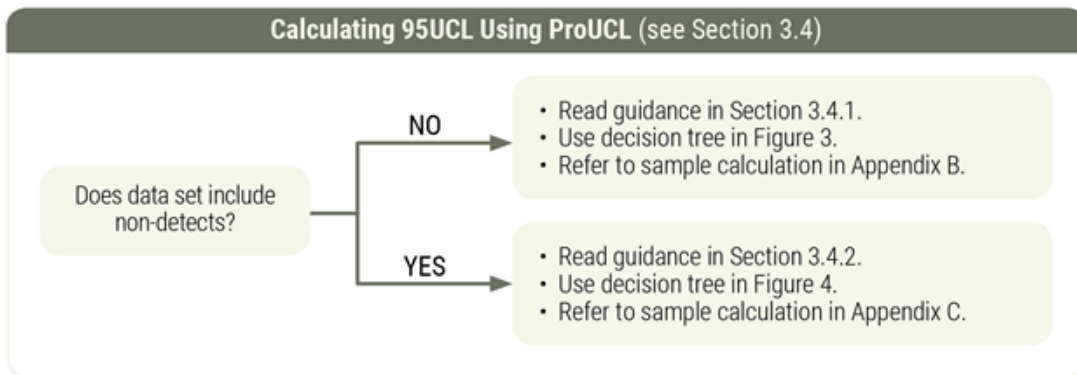Figure 1. Selection of EPC Statistics for Different Exposure Durations

Since intermediate and chronic duration exposures are of interest, the health assessor should use a 95UCL.
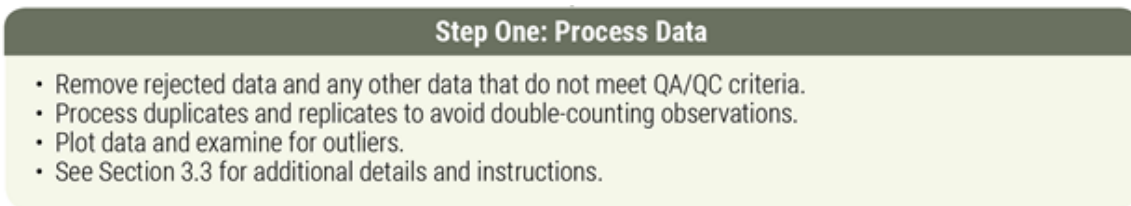
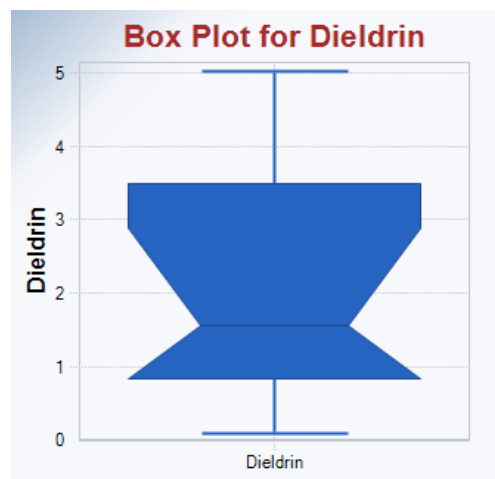2) Calculate the appropriate EPC identified in Question 1.

Answer:

Since these data do not contain non-detects and contain more than 8 samples, the health assessor should follow the decision tree in Figure 3 from ATSDR's guidance to calculate a 95UCL.

Calculating 95UCL Using ProUCL (see Section 3.4)

Does data set include non-detects?

NO
- Read guidance in Section 3.4.1.
- Use decision tree in Figure 3.
- Refer to sample calculation in Appendix B.

YES
- Read guidance in Section 3.4.2.
- Use decision tree in Figure 4.
- Refer to sample calculation in Appendix C.

Step 1: Process the Data

Step One: Process Data

- Remove rejected data and any other data that do not meet QA/QC criteria.
- Process duplicates and replicates to avoid double-counting observations.
- Plot data and examine for outliers.
- See Section 3.3 for additional details and instructions.

The data in Table 1 do not contain any rejected data, duplicates, or replicates. The data can be directly imported into ProUCL and examined for outliers with a boxplot as shown in the figure below. The boxplot does not indicate the presence of outliers (i.e., maximum concentration is only marginally higher than the mean).

Box Plot for Dieldrin

Step 2: Identify the Probability Distribution

Step Two: Calculate 95UCL

Option A: 8–19 Samples

Step 2A:
In ProUCL's Statistical Tests menu, select→Goodness-of-Fit Tests →Full (w/o NDs) →GOF Statistics. Select the distribution (normal, lognormal, or gamma) with correlation coefficient closest to 1.*

Step 2B:
In ProUCL's UCLs/EPCs menu, select →Full (w/o NDs) → Selected Distribution. Use the following values as EPCs depending on the selected distribution:
- Normal distribution – Use the 95% Student's t-UCL.
- Lognormal distribution – 95% Chebyshev UCL.**
- Gamma distribution – Use the 95% Adjusted Gamma UCL for n<50.

Option B: ≥ 20 Samples

- In ProUCL's UCLs/EPCs menu, select →Full (w/out NDs)→ Non-parametric. Use the 95% Percentile Bootstrap UCL as the EPC.

The highest correlation coefficient (0.965) is for a normal distribution as shown below:

Normal GOF Test Results

Correlation Coefficient R	0.965
Shapiro Wilk Test Statistic	0.916
Shapiro Wilk Critical (0.05) Value	0.842
Approximate Shapiro Wilk P Value	0.412
Lilliefors Test Statistic	0.246
Lilliefors Critical (0.05) Value	0.262

Data appear Normal at (0.05) Significance Level

Gamma GOF Test Results

Correlation Coefficient R	0.954
A-D Test Statistic	0.285
A-D Critical (0.05) Value	0.741
K-S Test Statistic	0.154
K-S Critical(0.05) Value	0.272

Data appear Gamma Distributed at (0.05) Significance Level

Lognormal GOF Test Results

Correlation Coefficient R	0.935
Shapiro Wilk Test Statistic	0.883
Shapiro Wilk Critical (0.05) Value	0.842
Approximate Shapiro Wilk P Value	0.117
Lilliefors Test Statistic	0.203
Lilliefors Critical (0.05) Value	0.262

Data appear Lognormal at (0.05) Significance Level

For exposure units having between 8 and 19 samples with no non-detects that are best fit by a normal distribution, ATSDR's guidance is for health assessors using ProUCL to select the 95% Student's t-UCL. In this example, the selected EPC is equal to 3.207 mg/kg as shown below:

Normal UCL Statistics for Uncensored Full Data Sets

User Selected Options	
Date/Time of Computation	ProUCL 5.16/12/2019 3:23:49 PM
From File	WorkSheet.xls
Full Precision	OFF
Confidence Coefficient	95%

Dieldrin

General Statistics

Total Number of Observations	10	Number of Distinct Observations	9
		Number of Missing Observations	0
Minimum	0.11	Mean	2.23
Maximum	5	Median	1.55
SD	1.686	SD of logged Data	1.143
Coefficient of Variation	0.756	Skewness	0.536

Normal GOF Test

Shapiro Wilk Test Statistic	0.916	**Shapiro Wilk GOF Test**	
5% Shapiro Wilk Critical Value	0.842	Data appear Normal at 5% Significance Level	
Lilliefors Test Statistic	0.246	**Lilliefors GOF Test**	
5% Lilliefors Critical Value	0.262	Data appear Normal at 5% Significance Level	
Data appear Normal at 5% Significance Level			

Assuming Normal Distribution

95% Normal UCL		95% UCLs (Adjusted for Skewness)	
95% Student's-t UCL	3.207	95% Adjusted-CLT UCL (Chen-1995)	3.203
		95% Modified-t UCL (Johnson-1978)	3.222

Suggested UCL to Use

95% Student's-t UCL	3.207	

Note: Suggestions regarding the selection of a 95% UCL are provided to help the user to select the most appropriate 95% UCL.
Recommendations are based upon data size, data distribution, and skewness.
These recommendations are based upon the results of the simulation studies summarized in Singh, Maichle, and Lee (2006).
However, simulations results will not cover all Real World data sets; for additional insight the user may want to consult a statistician.

Step 3: Perform a "Reality Check"

Step Three: Perform "Reality Check" on Results

- Compare 95UCL to maximum and arithmetic mean concentrations.
- Check for influence of data outliers.
- Refer to Section 3.7 for specific guidance.

The 95UCL of 3.207 mg/kg is less than the maximum concentration of 5.0 mg/kg and greater than the arithmetic mean of 2.23 mg/kg, as expected (and less than three times so). These comparisons confirm that the selected 95UCL is reasonable for use.

Case Study 2: Vinyl Chloride in Air

ATSDR received a petition from a state agency to assist in evaluating environmental health concerns of a residential community located near a manufacturing facility. The community members have previously complained of odors in the summer resembling that of volatile organic compounds (VOCs). ATSDR accepted the state's request to assist in evaluating their environmental monitoring data.

The nearest residences are in a housing complex 1000 feet from the plant boundary. A health assessor assigned the housing complex property as an exposure unit. An ambient air monitoring station was installed at the housing complex, and 24-hour air monitoring samples were collected every three days from June through August. For every fifth sampling event, a co-located sample was collected as a means of quality control. At this hypothetical site, vinyl chloride, a VOC, was identified as a contaminant of concern. Measurements are shown in Table 1. ATSDR minimal risk levels (MRLs) for vinyl chloride are available for acute, intermediate, and chronic-duration inhalation exposures. The health assessor has already determined that acute exposures do not pose a risk to health, and that intermediate-duration exposure are of concern because of the presence of odors during summer months only.

Table 1
Environmental Sampling Data for Vinyl Chloride [2]

Sample ID	Result ($\mu g/m^3$)	Sample ID	Result ($\mu g/m^3$)	Sample ID	Result ($\mu g/m^3$)
Sample 1a	0.42	Sample 11a	0.32	Sample 21a	0.15
Sample 1b	0.45	Sample 11b	0.51	Sample 21b	0.11
Sample 2	0.29	Sample 12	< 0.01	Sample 22	0.19
Sample 3	< 0.01	Sample 13	0.66	Sample 23	0.06
Sample 4	1.02	Sample 14	0.15	Sample 24	0.81
Sample 5	0.12	Sample 15	0.35	Sample 25	0.04
Sample 6a	7.57	Sample 16	0.35	Sample 26a	0.02
Sample 6b	9.10	Sample 16b	0.43	Sample 26b	0.08
Sample 7	0.45	Sample 17	< 0.01	Sample 27	0.79
Sample 8	0.89	Sample 18	0.06	Sample 28	0.09
Sample 9	0.07	Sample 19	0.36	Sample 29	0.61
Sample 10	< 0.01	Sample 20	0.04	Sample 30	1.52

Review the data provided above and answer the following questions.

[2] *NOTE: All sampling data provided are for a hypothetical site. Assume that the sampling data are of a known and high quality and are representative of actual exposure concentrations.*

Questions

1) What EPC statistic should the health assessor select for determining intermediate duration exposure at this site? Identify the criteria considered for making your selection.

Answer:

The environmental data have been screened for quality, the exposure unit is defined, and vinyl chloride has been identified as a contaminant of concern. Therefore, the health assessor should follow the flow chart in Figure 1 from the EPC guidance document.
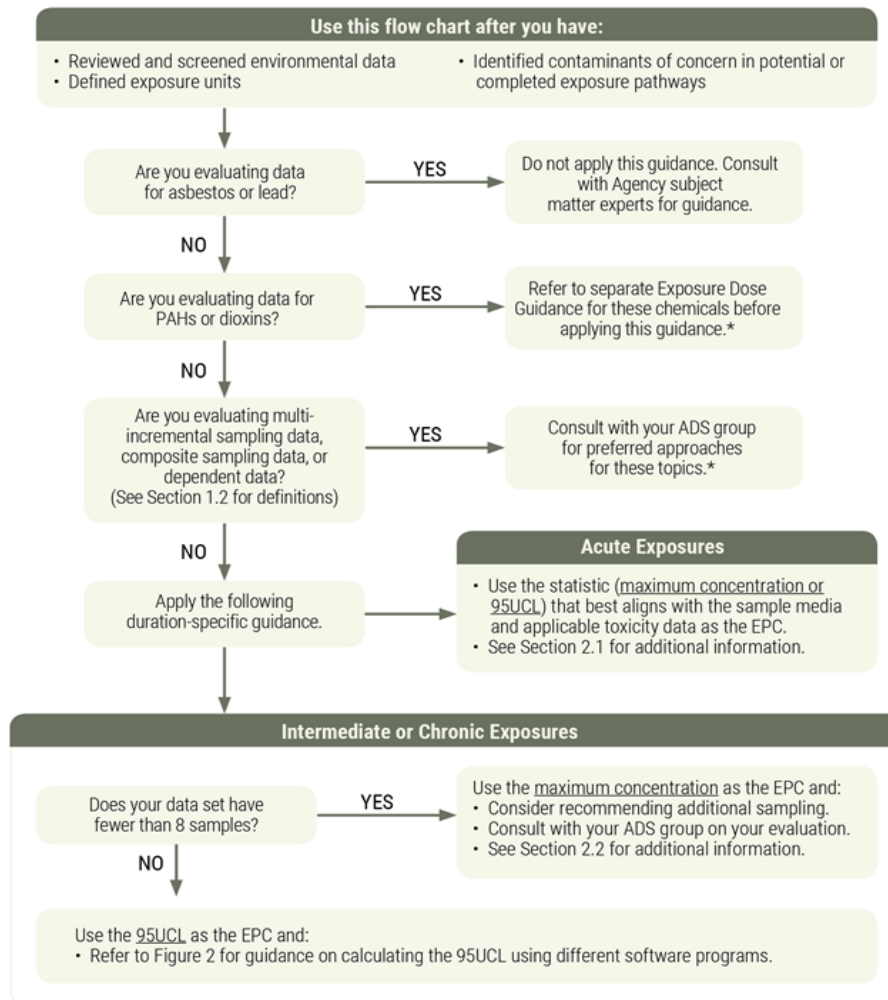
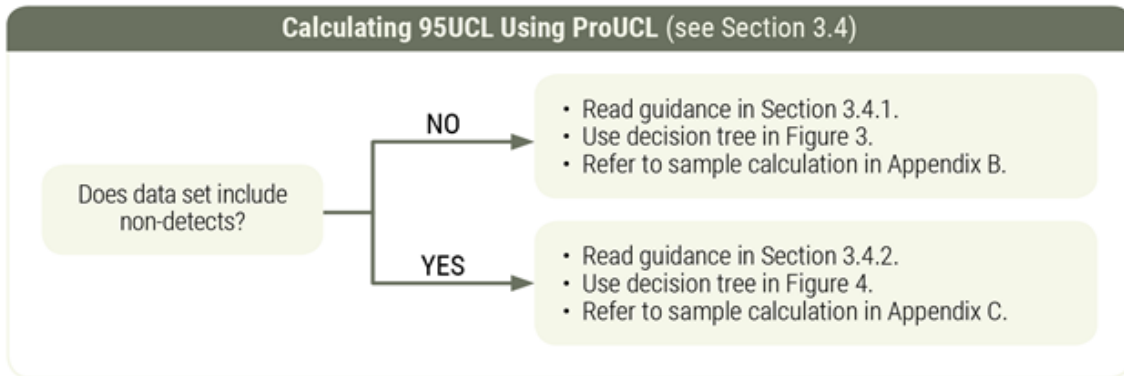Figure 1. Selection of EPC Statistics for Different Exposure Durations

Because intermediate duration exposures are of interest, the health assessor should use a 95UCL.

2) Calculate the appropriate EPC identified in Question 1.

Answer:
Since the data set contains non-detects and more than 8 samples, the health assessor should follow the decision tree in Figure 4 from ATSDR's guidance to calculate a 95UCL.

Calculating 95UCL Using ProUCL (see Section 3.4)

Does data set include non-detects?

NO
- Read guidance in Section 3.4.1.
- Use decision tree in Figure 3.
- Refer to sample calculation in Appendix B.

YES
- Read guidance in Section 3.4.2.
- Use decision tree in Figure 4.
- Refer to sample calculation in Appendix C.

Step 1: Process the Data
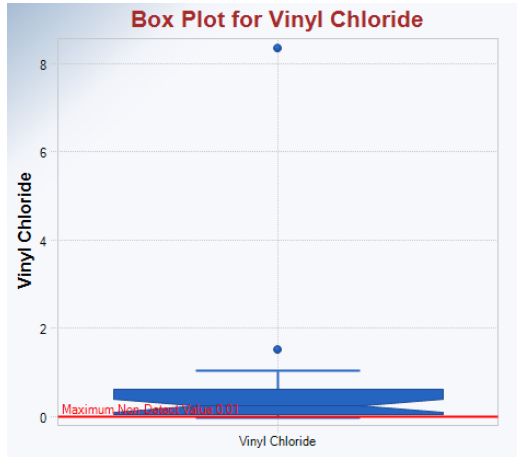
Step One: Process Data

- Remove rejected data and any other data that do not meet QA/QC criteria.
- Process duplicates and replicates to avoid double-counting observations.
- Plot data and examine for outliers.
- See Section 3.3 for additional details and instructions.

The data in Table 1 contain duplicates. These should be averaged such that every sampling event has a single concentration result, as shown here:
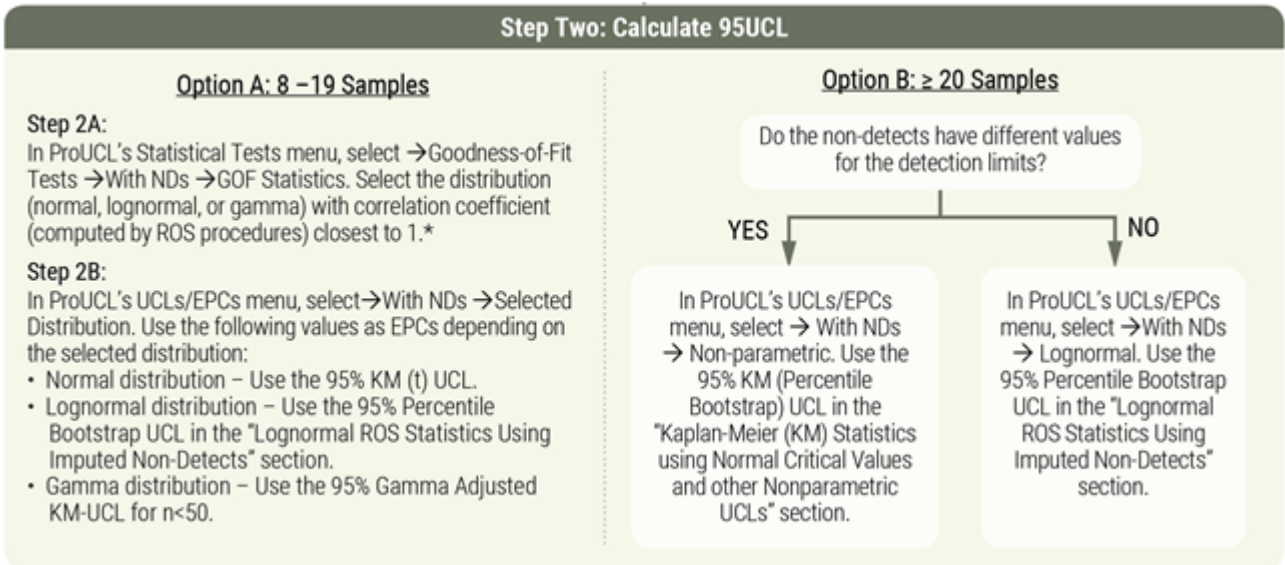
Sample	Results	Average
Sample 1a	0.42	0.435
Sample 1b	0.45	
Sample 6a	7.57	8.335
Sample 6b	9.10	
Sample 11a	0.32	0.415
Sample 11b	0.51	
Sample 16a	0.35	0.390
Sample 16b	0.43	
Sample 21a	0.15	0.130
Sample 21b	0.11	
Sample 26a	0.02	0.050
Sample 26b	0.08	

After calculating averages, import the data into ProUCL and examine for outliers with a boxplot as shown in the figure below. One concentration is notably higher than the other measurements, but it is still a valid result. Without a science-based reason to exclude them, this higher value should be kept in the dataset for 95UCL calculations.

Box Plot for Vinyl Chloride

Step 2: Calculate the 95UCL

Step Two: Calculate 95UCL

Option A: 8 –19 Samples

Step 2A:
In ProUCL's Statistical Tests menu, select →Goodness-of-Fit Tests →With NDs →GOF Statistics. Select the distribution (normal, lognormal, or gamma) with correlation coefficient (computed by ROS procedures) closest to 1.*

Step 2B:
In ProUCL's UCLs/EPCs menu, select→With NDs →Selected Distribution. Use the following values as EPCs depending on the selected distribution:
• Normal distribution – Use the 95% KM (t) UCL.
• Lognormal distribution – Use the 95% Percentile Bootstrap UCL in the "Lognormal ROS Statistics Using Imputed Non-Detects" section.
• Gamma distribution – Use the 95% Gamma Adjusted KM-UCL for n<50.

Option B: ≥ 20 Samples

Do the non-detects have different values for the detection limits?

YES | NO

In ProUCL's UCLs/EPCs menu, select → With NDs → Non-parametric. Use the 95% KM (Percentile Bootstrap) UCL in the "Kaplan-Meier (KM) Statistics using Normal Critical Values and other Nonparametric UCLs" section.

In ProUCL's UCLs/EPCs menu, select →With NDs → Lognormal. Use the 95% Percentile Bootstrap UCL in the "Lognormal ROS Statistics Using Imputed Non-Detects" section.

For exposure units having 20 or more samples and non-detect observations with the same value for the detection limits, ATSDR's guidance is for health assessors using ProUCL to select the 95% Percentile Bootstrap UCL. In this example, the selected EPC is equal to 1.144 µg/m^3 as shown below:

Lognormal GOF Test on Detected Observations Only			
Shapiro Wilk Test Statistic	0.951	**Shapiro Wilk GOF Test**	
5% Shapiro Wilk Critical Value	0.92	Detected Data appear Lognormal at 5% Significance Level	
Lilliefors Test Statistic	0.112	**Lilliefors GOF Test**	
5% Lilliefors Critical Value	0.17	Detected Data appear Lognormal at 5% Significance Level	
Detected Data appear Lognormal at 5% Significance Level			

Lognormal ROS Statistics Using Imputed Non-Detects			
Mean in Original Scale	0.613	Mean in Log Scale	-1.68
SD in Original Scale	1.504	SD in Log Scale	1.575
Geometric d_mean	0.186	95% Percentile Bootstrap UCL	1.144
95% t UCL (assumes normality of ROS data)	1.079	95% BCA Bootstrap UCL	1.452
95% Bootstrap t UCL	2.339	95% H-UCL (Log ROS)	1.681

Step 3: Perform a "Reality Check"

Step Three: Perform "Reality Check" on Results

- Compare 95UCL to maximum and arithmetic mean concentrations.
- Check for influence of data outliers.
- Refer to Section 3.7 for specific guidance.

The 95UCL of 1.144 $\mu g/m^3$ is less than the maximum concentration of 8.335 $\mu g/m^3$ and greater than the arithmetic mean of 0.613 $\mu g/m^3$ (see entry for "Mean in Original Scale"), as expected (and less than three times so). These comparisons confirm that the selected 95UCL is reasonable for use.

Because the 95UCL is less than three times the mean, the health assessor need not investigate influence of outliers.

Figure 1. Selection of EPC Statistics for Different Exposure Durations (Page 2)

The first box in this flow chart states to use this flow chart after you have: Reviewed and screened environmental data. Defined exposure units. Identified contaminants of concern in potential or completed exposure pathways.

An arrow points from this box to the next box which asks the question: Are you evaluating data for asbestos or lead? If the answer is yes, an arrow points to a new box which states: Do not apply this guidance. Consult with Agency subject matter experts for guidance.

If the answer is no, an arrow points to the next question: Are you evaluating data for PAHs or dioxins? If the answer is yes, an arrow points to a new box which states: Refer to separate Exposure Dose Guidance for these chemicals before applying this guidance. This is followed by a footnote which states: Guidance documents for these chemicals and topics are under development. Until they are issued, consult with your Branch ADS on preferred approaches.

If the answer is no, an arrow points to the next question: Are you evaluating multi-incremental sampling data, composite sampling data, dependent data, or times series? (See Section 1.2 for definitions). If the answer is yes, an arrow points to a new box which states: Consult with your Branch ADS for preferred approaches for these topics. This is followed by a footnote which states: Guidance documents for these chemicals and topics are under development. Until they are issued, consult with your Branch ADS on preferred approaches.

If the answer is no, an arrow points the next box which states: Apply the following duration specific guidance. This box has arrows pointing towards two new boxes. The first box is labeled Acute Exposures and states: Use the statistic (maximum concentration or 95UCL) that best aligns with the sample media and applicable toxicity data as the EPC. See Section 2.1 for additional information.

The second box is labeled Intermediate or Chronic Exposures and has a question box and two answers within it. The question box asks: Does your data have fewer than 8 samples? If the answer is yes, an arrow points to a new box which states: Use the maximum concentration as the EPC and: Consider recommending additional sampling. Consult with your Branch ADS on your evaluation. See Section 2.2 for additional information. If the answer is no, an arrow points to another new box which states: Use the 95UCL as the EPC and: Refer to Figure 2 for guidance on calculating the 95UCL using different software programs.

Calculating 95UCL Using ProUCL (Page 3)

Within this box, there is a question box which asks: Does data set include no-detects?

If the answer is no, an arrow points to a box which states: Read the guidance in Section 3.4.1. Use decision tree in Figure 3 of that guidance. Refer to sample calculation in Appendix B in that guidance.

If the answer is yes, an arrow points to a box which states: Read the guidance in Section 3.4.2. Use decision tree in Figure 4 of that guidance. Refer to sample calculation in Appendix C in that guidance.

Step 1: Process Data (Page 3)

This shows a screenshot of a flowchart from the EPC Guidance for Discrete Sampling. There is a single box labeled Step One: Process Data which states: Remove rejected data and any other data that do not meet QA/QC criteria. Process duplicates and replicates to avoid double-counting observations. Plot data and examine for outliers. See Section 3.3 for additional details and instructions.

Step Two: Calculate 95UCL (page 4)

The heading box is labeled Step Two: Calculate 95UCL and presents two options: Option A for 8-19 samples and Option B for greater than or equal to 20 samples. For Option A (8-19 samples) there are two steps stated: Step 2A: In ProUCL's Statistical Tests menu, select, Goodness-of-Fit Tests, Full (w/o NDs), GOF Statistics. Select the distribution (normal, lognormal, or gamma) with correlation coefficient closest to 1.

Step 2B: In ProUCL's UCLs/EPCs menu, select, Full (w/o NDs), Selected Distribution. Use the following values as EPCs depending on the selected distribution: Normal distribution – Use the 95% Student's t-UCL. Lognormal distribution – 95% Chebyshev UCL. Gamma distribution – Use the 95% Adjusted Gamma UCL for n<50.

For Option B for greater than or equal to 20 samples, there is one step stated: In ProUCL's UCLs/EPCs menu, select, Full (w/out NDs), Non-parametric. Use the 95% Percentile Bootstrap UCL as the EPC.

Step Two: Calculate 95UCL (page 10)

Step Two: Calculate 95 UCL presents two options: Option A for 8-19 samples and Option B for greater than or equal to 20 samples. For Option A (8-19 samples) there are two steps stated. Step 2A states: In ProUCL's Statistical Tests menu, select, Goodness-of-Fit Tests, With NDs, GOF Statistics. Select the distribution (normal, lognormal, or gamma) with correlation coefficient (computed by ROS procedures) closest to 1.

Step 2B states: In ProUCL's UCLs/EPCs menu, select, With NDs, Selected Distribution. Use the following values as EPCs depending on the selected distribution: Normal distribution – Use the 95% KM (t) UCL. Lognormal distribution – Use the 95% Percentile Bootstrap UCL in the "Lognormal ROS Statistics Using Imputed Non-Detects" section. Gamma distribution – Use the 95% Gamma Adjusted KM-UCL for n<50.

Option B for greater than or equal to 20 samples asks the following question: Do the non-detects have different values for detection limits? If the answer is yes, an arrow points to a box which states: In ProUCL's UCLs/EPCs menu, select, With NDs, Non-parametric. Use the 95% KM (Percentile Bootstrap) UCL in the "Kaplan-Meier (KM) Statistics using Normal Critical Values and other Nonparametric UCLs" section. If the answer is no, an arrow points to a box which states: In ProUCL's UCLs/EPCs menu, select, With NDs, Lognormal. Use the 95% Percentile Bootstrap UCL in the "Lognormal ROS Statistics Using Imputed Non-Detects" section.