GUIDANCE DOCUMENT

DIRECTIONS TO DETERMINE 95 PERCENT UPPER CONFIDENCE LEVEL OF THE MEAN USING USEPA PROUCL VERSION 4.0 SOFTWARE PACKAGE

June 2008

Table of Contents

I. ProUCL Version 4.00.02

A. Background

The determination of the 95% UCL of the mean shall be accomplished with a program that can properly evaluate non-detect results if needed. The use of the most current version of the EPA program ProUCL (ProUCL; currently Version 4.00.02) is recommended by the Department for the evaluation.

The EPA maintains the statistical program ProUCL. This statistical software program can be used to compute upper limits including the 95% UCL. The ProUCL software package provides several statistical methods that can be used on data sets with non-detect observations potentially having multiple detection limits (DLs). Information about the program and download directions can be found at http://www.epa.gov/nerlesd1/tsc/software.htm (for Department personnel, directions for downloading and installing the program will be provided by email). This program is available to the public at no cost to the user.

B. Instructions for use of the program

ProUCL provides an online help file that explains the program and how to use it. Basic directions for entering data and calculating the 95% UCL of the mean are outlined in Appendix D of this document.

C. Minimum number of distinct sample values and total sample points to be used in ProUCL 4

Based on its evaluation of the ProUCL program, the Department has identified that a minimum of three distinct sample values must be present in the data set in order for the program to function properly. In addition, because the calculated result is more robust with more samples, the Department is establishing that the minimum number of samples to be inputted into the program is seven data points. In those cases where there are two or fewer distinct sample values or six or fewer total sample points, compliance shall be determined by calculating the average for the data set.

D. Evaluation of ProUCL results

ProUCL will identify the recommended 95% UCL of the mean for the data set under evaluation. The result is identified as the "Potential UCL to Use" on the output table. This is the 95% UCL of the mean that should be used for evaluation of the inhalation exposure pathway. An example of an output table from ProUCL is provided as Appendix A of this document.

The Department has identified two circumstances in which the "Potential UCL to Use" requires further explanation.

- If ProUCL identifies more than one potential UCL to use, the Department has determined that the lower value shall be used in the evaluation of the inhalation exposure pathway.

- If the calculated 95% UCL of the mean is greater than all values in the data set, the Department has determined that the maximum sample value in the data set shall be used for evaluation of the inhalation exposure pathway.

II.   Instructions for using ProUCL version 4.00.02

1. Open ProUCL 4

2. The first step is to name the data column with a description that summarizes the data set. To name the column, left click on the gray box over the column number. This will select the column and it will change color to blue. Next, right click on the box over the column number. A new box appears with the title header name. Left click on this box and inset the data set name, then click ok.

3. If your data set includes non-detect values (ND values), name the adjacent column to the right in the same manner with "d_" or "D_" (the underscore is required) followed by exactly the same data set name. This column will differentiate between detected and ND values results and must start with the "d_" or "D_". Note that the program will not recognize that these two columns are associated if the data set names are different.

If your data set does not include ND values, it is not necessary to include the second column.

4. Begin entering the data in the column below the name. Each row is numbered for data entry. For non-detect data enter the detection limit. This shall include data assigned the "U" laboratory qualifier ("undetected").

5. In the adjacent column under the "d_" or "D_" header name, enter "1" for a detected concentration and enter "0" for a non-detected concentration. This column should only contain the values "0" or "1".

6. When all data have been entered, left click on UCL at the top of the page. A box will appear with the titles "Full" and "With NDs". From this box, left click on "Full" for a data set with all detected concentrations or left click on "With NDs" for a data set with non-detected concentrations.

7. A new box will appear with five (5) options. Left click on option "All".

8. A "Select variables" box will appear. In this box, left click on the data set you would like to evaluate. Data sets will be located on the left side of the screen under "Variables". Move the data set(s) to be evaluated to the right hand side of the box under the heading "Selected" by left clicking on the arrows. The next step is to left click on OK after all data sets for evaluation have been moved to the right side of the screen.

9. At this point, the data set(s) will be evaluated and the results shown on the screen.

10. ProUCL evaluates all results and selects the appropriate 95% UCL of the mean at the bottom of the results set under the heading "Potential UCL to Use". This is the 95% UCL of the mean that shall be used for your data set.

APPENDIX A: Example of ProUCL 4 Output Table

General UCL Statistics for Data Sets with Non-Detects	
User Selected Options	
From File	E:\Input.wst
Full Precision	OFF
Confidence Coefficient	95%
Number of Bootstrap Operations	2000
X	

General Statistics			
Number of Valid Data	18	Number of Detected Data	16
Number of Distinct Detected Data	16	Number of Non-Detect Data	2
		Percent Non-Detects	11.11%

Raw Statistics		**Log-transformed Statistics**	
Minimum Detected	0.0428	Minimum Detected	-3.151
Maximum Detected	287	Maximum Detected	5.659
Mean of Detected	40.54	Mean of Detected	1.601
SD of Detected	72.61	SD of Detected	2.841
Minimum Non-Detect	0.01	Minimum Non-Detect	-4.605
Maximum Non-Detect	0.01	Maximum Non-Detect	-4.605

UCL Statistics			
Normal Distribution Test with Detected Values Only		**Lognormal Distribution Test with Detected Values Only**	
Shapiro Wilk Test Statistic	0.595	Shapiro Wilk Test Statistic	0.902
5% Shapiro Wilk Critical Value	0.887	5% Shapiro Wilk Critical Value	0.887
Data not Normal at 5% Significance Level		Data appear Lognormal at 5% Significance	

	Level		
Assuming Normal Distribution		**Assuming Lognormal Distribution**	
DL/2 Substitution Method		DL/2 Substitution Method	
Mean	36.03	Mean	0.834
SD	69.46	SD	3.479
95% DL/2 (t) UCL	64.51	95% H-Stat (DL/2) UCL	214724
Maximum Likelihood Estimate(MLE) Method		Log ROS Method	
Mean	30.65	Mean in Log Scale	0.879
SD	73.1	SD in Log Scale	3.403
95% MLE (t) UCL	60.62	Mean in Original Scale	36.04
95% MLE (Tiku) UCL	58.76	SD in Original Scale	69.46
		95% Percentile Bootstrap UCL	65.52
		95% BCA Bootstrap UCL	77.72
Gamma Distribution Test with Detected Values Only		**Data Distribution Test with Detected Values Only**	
k star (bias corrected)	0.305	**Data appear Gamma Distributed at 5% Significance Level**	
Theta Star	132.9		
nu star	9.757		
A-D Test Statistic	0.535	**Nonparametric Statistics**	
5% A-D Critical Value	0.833	Kaplan-Meier (KM) Method	
K-S Test Statistic	0.833	Mean	36.04
5% K-S Critical Value	0.232	SD	67.5
Data appear Gamma Distributed at 5% Significance Level		SE of Mean	16.43
		95% KM (t) UCL	64.62
Assuming Gamma Distribution		95% KM (z) UCL	63.07
Gamma ROS Statistics using Extrapolated Data		95% KM (jackknife) UCL	64.52

Minimum	1E-09	95% KM (bootstrap t) UCL	115.9
Maximum	287	95% KM (BCA) UCL	65.07
Mean	36.03	95% KM (Percentile Bootstrap) UCL	65.28
Median	9.72	95% KM (Chebyshev) UCL	107.7
SD	69.46	97.5% KM (Chebyshev) UCL	138.7
k star	0.178	99% KM (Chebyshev) UCL	199.5
Theta star	202.5		
Nu star	6.407	**Potential UCLs to Use**	
AppChi2	1.851	95% KM (Chebyshev) UCL	107.7
95% Gamma Approximate UCL	124.7		
95% Adjusted Gamma UCL	141.9		
Note: DL/2 is not a recommended method.			