

Yolande Norman

From: Myron Fliegel
Sent: Friday, November 21, 2008 9:10 AM
To: Yolande Norman
Subject: FW: EPA comments on the statistical objectives
Attachments: EPA Comments on Statistical Objectives 111908.doc

-----Original Message-----

From: Purcell.Mark@epamail.epa.gov [mailto:Purcell.Mark@epamail.epa.gov]
Sent: Wednesday, November 19, 2008 1:01 PM
To: mark.jancin@veoliawater.com; Blickwedel, Roy (Corporate); larry.bush@ae.ge.com
Cc: tsaliassociates@hotmail.com; david.mayerson@state.nm.us; Diana Malone; Myron Fliegel
Subject: EPA comments on the statistical objectives

Larry, Roy, Mark,

Our statistician, Dennis Beal, completed his review of the two statistical objective packages for UNC.

I have attached EPA comments.,

Please let us know when your ready and we can schedule a call to discuss if needed.

Mark

Mark D. Purcell
Superfund Division (6SF-RL)
USEPA - Region 6
1445 Ross Avenue, Suite 1200
Dallas, TX 75202-2733
T 214.665.6707
F 214.665.6660
purcell.mark@epa.gov

(See attached file: EPA Comments on Statistical Objectives 111908.doc)

EPA COMMENTS

On the

United Nuclear Corporation's
Revised Church Rock Statistics – Objectives 1 and 2

Revised Church Rock Statistic Objective - Package 1

In general, this statistical approach is much improved from the previous approach. EPA has the following minor comments and questions.

1. On page 6, the text states that a nondetect represents a result less than its reporting limit (RL). Did UNC consider using the method detection limit (MDL) or sample quantitation limit (SQL) instead of the RL for nondetects? Sometimes, it is preferred to use the SQL over the RL, since the SQL incorporates matrix effects of the individual sample.
2. On page 6, the text states that the sign test was run on chemicals with 100% nondetects. Why were any statistical tests run on data with all nondetects? This was discussed before during a teleconference between EPA and UNC. You will get strange results from statistical tests if you are using all nondetects. No statistical test should be run with all nondetects.
3. The text does a good job of explaining why single hypothesis tests (sign test and WSR) do not work well for skewed data and data with large fraction of nondetects. EPA agrees that using ProUCL to calculate the UCL95 using Kaplan-Meier for data with nondetects and compare UCL95 with comparison values is better than the single hypothesis tests, although some guidance argues that UCL95 should not be compared to single values, only to other averages since the UCL95 is an upper bound on an average.
4. The conclusion of using two-sample tests to compare compliance data with background data is acceptable.

Revised Church Rock Statistic Objective – Package 2

1. Please provide some censored or normal probability plots for these constituents to accompany the detailed ProUCL output tables where detects have different symbols than non-detects.
2. On page 4, last paragraph, it states there are 6 constituents in Zone 1 (Table 3). The text should say there are 7 according to Tables 3 and 5.
3. The tables in the appendix need individual table numbers (e.g., A.1, A.2, A.3) so they can be referred to specifically.
4. The appendix table for the Southwest Alluvium has the variance and standard deviation values for Mn, NO₃_as_N and U switched. The recommended UCL95 values for Mn and NO₃_as_N are really UCL97.5 values. The table should footnote these values so the reader knows the recommended UCL95s are really UCL97.5 values.
5. Some cells of the appendix tables are missing when they should not be. For example, the Southwest Alluvium table number of non-detects row is missing for Lab_TDS column. Also, Be and Cd columns have min and max detected cells blank since there are no detects, but the other constituents with no detects have “N/A” in these cells. These should be consistent.
6. Similarly, cells of the Zone 1 appendix table for number of non-detects for SO₄ and NO₃_as_N are blank.
7. ProUCL chose the UCL99 for NO₃_as_N for Zone 1 in appendix table because the standard deviation of the log-transformed data was 1.037, which fell into the range of 1 to 2 according to Table 2.5 in the ProUCL technical guidance document. The UCL99 = 243.6, which far exceeds the maximum result of 200. Note that Table 2.5 of the ProUCL technical guidance document would have recommended the 95% Chebyshev UCL if the standard deviation of the log-transformed data ranged from 0.5 to 1. Since the UCL99 exceeds the range of the data and 1.037 barely falls into the 1 to 2 recommended range, it would have been more appropriate to choose the 95% UCL instead of the 99% UCL. The UCL99 is way too conservative. This is an example where the users have to use experienced judgment when interpreting the ProUCL recommendations for their particular data set. If 243.6 is used, then it should be footnoted it is really a UCL99 and not a UCL95.
8. Table 3 of the report recommends a UCL of 1.95 for Mn, but in the appendix Zone 1 table, a larger H-UCL95 of 2.02 from Land's is reported. Since the data are lognormally distributed, why wasn't Land's H-UCL95 = 2.02 chosen? The text of the report stated the larger of two recommended UCLs would be chosen, but it wasn't in this case.

9. The UCL values in Table 4 for Al, As and NO3_as_N should be footnoted that they are UCL97.5 values, not UCL95.
10. The appendix table for Zone 3 has no distribution stated for V. Also, Pb in this table has no min or max non-detected data in their respective cells even though the data are all non-detects.
11. Since some copy/paste errors were previously noted in the appendix tables from the ProUCL output, these tables need to be rechecked for accuracy.

EPA COMMENTS

On the

United Nuclear Corporation's
Revised Church Rock Statistics – Objectives 1 and 2

Revised Church Rock Statistic Objective - Package 1

In general, this statistical approach is much improved from the previous approach. EPA has the following minor comments and questions.

1. On page 6, the text states that a nondetect represents a result less than its reporting limit (RL). Did UNC consider using the method detection limit (MDL) or sample quantitation limit (SQL) instead of the RL for nondetects? Sometimes, it is preferred to use the SQL over the RL, since the SQL incorporates matrix effects of the individual sample.
2. On page 6, the text states that the sign test was run on chemicals with 100% nondetects. Why were any statistical tests run on data with all nondetects? This was discussed before during a teleconference between EPA and UNC. You will get strange results from statistical tests if you are using all nondetects. No statistical test should be run with all nondetects.
3. The text does a good job of explaining why single hypothesis tests (sign test and WSR) do not work well for skewed data and data with large fraction of nondetects. EPA agrees that using ProUCL to calculate the UCL95 using Kaplan-Meier for data with nondetects and compare UCL95 with comparison values is better than the single hypothesis tests, although some guidance argues that UCL95 should not be compared to single values, only to other averages since the UCL95 is an upper bound on an average.
4. The conclusion of using two-sample tests to compare compliance data with background data is acceptable.

Revised Church Rock Statistic Objective – Package 2

1. Please provide some censored or normal probability plots for these constituents to accompany the detailed ProUCL output tables where detects have different symbols than non-detects.
2. On page 4, last paragraph, it states there are 6 constituents in Zone 1 (Table 3). The text should say there are 7 according to Tables 3 and 5.
3. The tables in the appendix need individual table numbers (e.g., A.1, A.2, A.3) so they can be referred to specifically.
4. The appendix table for the Southwest Alluvium has the variance and standard deviation values for Mn, NO₃_as_N and U switched. The recommended UCL95 values for Mn and NO₃_as_N are really UCL97.5 values. The table should footnote these values so the reader knows the recommended UCL95s are really UCL97.5 values.
5. Some cells of the appendix tables are missing when they should not be. For example, the Southwest Alluvium table number of non-detects row is missing for Lab_TDS column. Also, Be and Cd columns have min and max detected cells blank since there are no detects, but the other constituents with no detects have “N/A” in these cells. These should be consistent.
6. Similarly, cells of the Zone 1 appendix table for number of non-detects for SO₄ and NO₃_as_N are blank.
7. ProUCL chose the UCL99 for NO₃_as_N for Zone 1 in appendix table because the standard deviation of the log-transformed data was 1.037, which fell into the range of 1 to 2 according to Table 2.5 in the ProUCL technical guidance document. The UCL99 = 243.6, which far exceeds the maximum result of 200. Note that Table 2.5 of the ProUCL technical guidance document would have recommended the 95% Chebyshev UCL if the standard deviation of the log-transformed data ranged from 0.5 to 1. Since the UCL99 exceeds the range of the data and 1.037 barely falls into the 1 to 2 recommended range, it would have been more appropriate to choose the 95% UCL instead of the 99% UCL. The UCL99 is way too conservative. This is an example where the users have to use experienced judgment when interpreting the ProUCL recommendations for their particular data set. If 243.6 is used, then it should be footnoted it is really a UCL99 and not a UCL95.
8. Table 3 of the report recommends a UCL of 1.95 for Mn, but in the appendix Zone 1 table, a larger H-UCL95 of 2.02 from Land's is reported. Since the data are lognormally distributed, why wasn't Land's H-UCL95 = 2.02 chosen? The text of the report stated the larger of two recommended UCLs would be chosen, but it wasn't in this case.

9. The UCL values in Table 4 for Al, As and NO3_as_N should be footnoted that they are UCL97.5 values, not UCL95.
10. The appendix table for Zone 3 has no distribution stated for V. Also, Pb in this table has no min or max non-detected data in their respective cells even though the data are all non-detects.
11. Since some copy/paste errors were previously noted in the appendix tables from the ProUCL output, these tables need to be rechecked for accuracy.