WRG

GUIDA, SLAVICH & FLORES
A PROFESSIONAL CORPORATION
The Environmental Law Firm℠

Dallas • Austin

Attorneys and Counselors
750 N. ST. PAUL STREET, SUITE 200
DALLAS, TEXAS 75201

TEL - 214.692.0009
FAX - 214.692.6610
www.guidaslavichflores.com

ENTERED

RECEIVED
JUL 2011
Hazardous Waste Bureau

July 15, 2011

Mark Hansen (6EN-H) *(via U.S. Mail)*
Associate Director, Hazardous Waste Enforcement Branch
U.S. EPA – Region 6
1445 Ross Avenue, Suite 1200
Dallas, TX 75202-2733

Jeffrey Clay (6RC-ER) *(via email and U.S. Mail)*
RCRA Legal Branch
Office of Regional Counsel
U.S. EPA – Region 6
1445 Ross Avenue, Suite 1200
Dallas, TX 75202-2733

John Kieling *(via U.S. Mail)*
Acting Chief, Hazardous Waste Bureau
New Mexico Environment Department
2905 Rodeo Park Drive East, Building 1
Santa Fe, NM 87505-6303

Re: Western Refining Southwest, Inc. Gallup Refinery ("Western")
Consent Agreement and Final Order,
No. RCRA-06-2009-0936, Effective August 26, 2009 (the "CAFO")

Dear Sirs:

Pursuant to the above CAFO, Western submits for consideration the following proposed monitoring plans: (i) a Toxicity Characteristic Monitoring Plan, Interim Measures Phase; and (ii) a Toxicity Characteristic Monitoring Plan, Upgraded Wastewater Treatment System. These plans are proposed to establish precise protocols

Austin Office

816 CONGRESS AVENUE, SUITE 1500
AUSTIN, TEXAS 78701
TEL - 512.476.6300
FAX - 512.476.6331

for demonstrating compliance under the CAFO that are consistent with, and more stringent than, the federal and State of New Mexico hazardous waste programs and representative sampling requirements. The first plan is proposed to be effective until the completion of Milestone 9 (the final milestone) under Paragraph 100.I of the CAFO, i.e. during the Interim Measures period. The Final Plan would then become effective and remain in place until the termination of the CAFO. Note that these plans are written as though they were approved final plans, but are draft proposals.

Western welcomes and seeks open discussions with the agencies about these proposed plans. In order to provide an opportunity for the agencies to begin providing initial input, questions, comments, concerns etc. without delay and to hear any additional explanation from Western that would be useful, Western would like to propose a conference call with the EPA and NMED plan reviewers (and anyone else who the agencies would like to include) on either July 20, 2011 at 1:30 p.m. MDT (2:30 p.m. CDT), or July 21, 2011 at 9:30 a.m. MDT (10:30 a.m. CDT). If these times are not convenient, we are, of course, glad to reschedule at a time when the agencies are available and ask that you suggest some alternate times. In the interest of time, we are emailing a copy of this letter and attached plans to Jeffrey Clay of EPA and asking that he distribute them to the appropriate EPA and NMED representatives. We will coordinate with Mr. Clay to confirm everyone's availability, and will circulate a conference number when a call time is finalized. We hope that this call will simply be the first in a series of conversations in which the parties share their thoughts and suggestions freely.

Western appreciates the commitment of the agencies in considering these monitoring plans, and looks forward to working with the agencies in the future.

Sincerely,

Jean Flores

Jean M. Flores
Direct Dial: 214.692.0017
E-Mail: flores@guidaslavichflores.com

Toxicity Characteristic Monitoring Plan Interim Measures Phase (Air Stripper)

Prepared for

Western Refining Southwest, Inc.
Gallup, New Mexico

by

Tischler/Kocurek
Registration No. F-3277
Round Rock, Texas

717
Version 1.0

July 2011

DRAFT

Western Refining Southwest Gallup Refinery
Interim Measures Phase (Air Stripper)
TC Characteristic Monitoring Plan

Introduction

This monitoring plan describes the procedures that are followed by the Western Refining Southwest, Inc. (Western) Gallup Refinery to monitor influent wastewater to Aerated Lagoon-1 (AL-1) for evaluation of the Toxicity Characteristic (TC) for hazardous wastes.[1] Monitoring of the wastewater in accordance with this plan provides documentation that the plant wastewater (process wastewater and storm water from process areas), prior to entering the most upstream surface impoundment (AL-1), is non-hazardous with respect to the TC.

This plan describes the toxicity characteristic, the regulatory basis for the monitoring approach, and how the monitoring data will be analysed to demonstrate continuing compliance. It also describes the results of the 16 months of monitoring between January 10, 2010 and June 30, 2011.

The Toxicity Characteristic

RCRA Definition

The Toxicity Characteristic is one criterion by which a waste is defined as hazardous under the Resource Conservation and Recovery Act (RCRA). TC limits have been established for 40 compounds as shown in Table 1 at 40 CFR 261.24(b). If a representative sample of a waste equals or exceeds the TC limit, it is considered a characteristic hazardous waste and is subject to the applicable sections of the hazardous waste regulations.

Although the hazardous waste regulations do not specify the time period over which a representative sample must be calculated for a continuous discharge such as plant wastewater, TC limits were developed by the United States Environmental Protection Agency (EPA) based on a long-term exposure to contaminated ground water. The ground water model used by EPA assumes a chemical that leaches into the ground water is supplied by a constant mass flux from an infinite landfill or impoundment source (55 Fed. Reg. 11817) and that an individual's exposure over a lifetime is constant (55 Fed. Reg. 11823). Therefore, EPA assumed that an individual would be subject to continuous chronic exposure to a chemical by drinking contaminated water over the individual's lifetime (55

[1] 40 CFR (Code of Federal Regulations) 261.24

DRAFT

Fed. Reg. 11813). Accordingly, EPA developed long-term exposure limits for the TC chemicals and promulgated these in Table 1 of Part 261.

Representative Sample

In order to determine if a waste is TC hazardous, a representative sample of the waste must be compared to the TC limit (40 CFR 261.24(a)). A representative sample is defined at 40 CFR 260.10 as "a sample of a universe or whole (e.g., waste pile, lagoon, ground water) which can be expected to exhibit the average properties of the universe or whole." Regulations issued by the New Mexico Environmental Department (NMED) (20.4.1.100, 20.4.1.200 NMAC) incorporate the EPA regulations cited above by reference. Chapter 9 of EPA's *Test Methods for Evaluating Solid Waste* (SW-846) is the Agency's guidance for collecting representative samples.

EPA acknowledges in this definition that average[2] waste concentrations and not individual grab or composite samples must be used when determining if a waste is TC hazardous. This position is emphasized in Chapter 9 of SW-846, which states that: "it is not the variation of individual sample concentrations that is of ultimate concern, but rather the variation that characterizes $\bar{x}$ itself" ($\bar{x}$ being the average).

Other EPA documents provide additional regulatory guidance on representative sampling for hazardous waste characterization. In these, EPA points out the need to collect a representative sample, which may be developed from multiple grab samples, and using statistics to incorporate a level of confidence in the average.[3]

Chapter 9 of SW-846 provides guidance on the number of waste samples that are appropriate for obtaining a representative sample (Table 9-1) when data are available (i.e., when testing is used to determine if a waste is hazardous) to characterize the probability distribution of the waste characteristic to be measured. Equation 8 in Table 9-1 is:

$$n = \frac{t_{0.20}^2}{\Delta^2} \times s^2$$

[2] The term "average properties" as used in the definition of a representative sample is the statistical measure of central tendency of the sample distribution that is representative of the population ("the universe or whole"). If the sample distribution of the data is normal, the arithmetic mean is the measure of central tendency. If the sample distribution of data is lognormal, the geometric mean (median) is the measure of central tendency. For distributions that are not normal or lognormal, the median or mode will be the measure of central tendency depending on the characteristics of the distribution.

[3] Letter from Mike Petruska, EPA Regulatory Development Branch, to Michael Roberts, Maintenance Engineering, June 22, 1995. EPA Memorandum, "Characterization of Nitrocellulose Filter Fabric Waste," from Elizabeth A. Cotsworth, EPA Acting Director, Office of Solid Waste to Barbara A. Finazzo, EPA Region 2 Director, Division of Environmental Science and Assessment, February 11, 1998. EPA, *Guidance for the Sampling and Analysis of Municipal Waste Combustion Ash for the Toxicity Characteristic*, EPA 530-R-95-036, Office of Solid Waste, Washington, D.C., June 1995.

DRAFT

where: n = appropriate number of samples
s = standard deviation of sample measurements
t = Student's t at 0.20 probability
$\Delta = R - \bar{x}$
$\bar{x}$ = mean of sample measurements
R = 0.5 mg/L (TC threshold for benzene)

Air stripper effluent data collected by Western from January 10, 2010 through June 30, 2011 were used in Equation 8 to calculate the number of effluent samples required to represent a representative sample as the term is used at 40 CFR 261.24(a).

$\bar{x}$ = 0.20 mg/L
s = 1.0936 mg/L
$t_{0.20}$ = 0.8416 (424 degrees of freedom)
$\Delta = 0.5 - 0.2 = 0.30$ mg/L
n = 9 samples

Under the interim measures plan, Western samples the stripper effluent as frequently as once/day. Therefore, a 9-sample moving average is appropriate to define a representative sample of treated effluent to determine if a discharge achieves the benzene threshold for TC. The interim measures plan does allow the sampling frequency to be reduced provided that Western demonstrates a period of compliance and NMED agrees to the reduced frequency.

Statistical Calculations

This section describes the statistical calculations for the rolling 9-sample average and how this average will be used to determine compliance with the TC concentration. A simplified flow sheet of the process is provided in Figure 1.

Fitting the Data to a Distribution

The first step in evaluating any data set is to determine what type of data distribution best fits the data because the type of distribution will affect the choice of statistics. With most data, the goal is to use normal distribution statistics. This is because there are many statistical tests for normal distributions and the calculations are relatively simple.

If the original data do not fit a normal distribution, the data can be transformed by any number of functions to produce a normal data set. The most common transformation function for environmental data is the natural-log. So often do environmental data follow the lognormal distribution, that EPA recommends the distribution for daily values unless the distribution clearly does not fit the data (*Technical Support Document for Water Quality-Based Toxics Control,* (TSD),

DRAFT

Appendix E). Two other data transformation methods recommended by EPA in SW-846 are the square root (when the mean is approximately equal to the variance) and the arcsine (when the mean is less than the variance).

The simplest, and often most effective way to determine goodness-of-fit, is to graph the data as recommended by EPA in its TSD (Appendix E) and in SW-846 (Chapter 9). If the data are normally distributed, they should plot as a straight line on a probability graph. If the data fit a lognormal distribution then the logarithms of the concentrations will be a straight line on a probability graph.

If the graphical probability analysis indicates that the sample data deviate significantly from a straight line on a probability graph, then a statistical test will be used to evaluate the goodness-of-fit to determine the appropriate distribution to be used for the calculation of the sample mean and variance. The Shapiro-Wilk test will be used to evaluate the goodness-of-fit for the normal and lognormal probability distributions when the visual comparison shows a significant deviation from normality. In such cases, the distribution that has the best fit to the sample data, as determined by the Shapiro-Wilk Test, will be used to calculate the mean and variance of the database for the confidence interval calculation.

A normal distribution is assumed for the data generated by this sampling plan. This assumption will be checked as data are generated. If the normal distribution is found to be unsuitable, the lognormal distribution will tested as a substitute. If neither the normal or lognormal distributions fit the database, a better fitting distribution will be selected.

Managing Censored Data

Concentrations that are reported as below the quantitation level of the analytical method for benzene are called *censored* values in statistical analyses. The simplest method for dealing with censored samples is to substitute one-half the reported quantitation (or detection) limit for each of the censored values in the database. The substitution method assumes that the concentrations below the quantitation limit are normally distributed and that the median/mean is the best estimate of the average concentration of the censored values in a data set. If the substitution method is used for databases with a large number of censored observations it will result in biased estimates of the mean and variance of the population. When a database contains a significant fraction of censored samples, a more robust statistical method for dealing with censored data is required. There are several such methods described in the scientific literature,[4] and a commonly used method is Cohen's maximum likelihood estimator (MLE) method. Cohen's MLE method has been shown by a number of studies to provide acceptable

[4] Berthouex, P.M. and Brown, L.C., 1994, *Statistics for Environmental Engineers*, CRC Press Boca Raton.

DRAFT

estimates of the mean and variance of databases with a significant amount of censoring.[5]

If there are censored samples in the 9-sample database, then Cohen's MLE method will be used to estimate the mean and variance of the sample population. The resulting estimated sample mean and variance will be used as described below to calculate the confidence interval that is used to assess compliance.

Descriptive Statistics

Equations from SW-846, as shown below, will be used to calculate the mean, the variance, and the standard deviation of the analytical data. According to SW-846, the equations shown are for simple random sampling as well as systematic random sampling. The monitoring data for the Western waste characterization are characterized as systematic random sampling, which according to SW-846, is when the first sample is randomly selected and all succeeding samples are taken at fixed time intervals. Because the hydraulic retention time in the collection system and treatment system (equalization tank, API separator, air strippers) are short, essentially any sampling frequency greater than a few hours will achieve the requirement for random samples.

Sample Mean (arithmetic average)

Equation (2a), SW-846

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

where

$\bar{x}$ = sample mean
x_i = individual data values, i=1…n[6]
n = number of data points

When the mean is calculated from data that have been transformed by the natural log function, it must be translated back to arithmetic scale by the exponential function as shown below. When the database fits a lognormal distribution, the mean that is calculated is the geometric mean (median) of the sample database, which is the measure of central tendency of a lognormal distribution.[7]

Translation of Log-Mean to Arithmetic Scale

Equation (A)

$$\bar{y} = e^{\bar{x}}$$

[5] Ibid.
[6] These data values will be transformed if necessary to meet the normal probability distribution assumption.
[7] Op cit., Berthouex, P.M. and Brown, L.C.

DRAFT

where:

$\bar{y}$ = sample geometric mean translated to arithmetic scale
e = the exponential function, approximately 2.718281828

Sample Variance

Equation (3a), SW-846

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}$$

where:

s^2 = variance

Sample Standard Deviation

Equation (4), SW-846

$$s = \sqrt{s^2}$$

where:

s = standard deviation

Upper Confidence Limit

In Chapter 9 of SW-846, EPA suggests a method for adding a degree of conservatism to any calculated average when determining compliance with TC limits. EPA suggests that the upper limit of an 80% confidence interval[8] be calculated for the average and then compared with the TC limit. The equation in SW-846 for the 80% confidence interval is given below:

Equation (6), SW-846

$$CI = \bar{x} \pm t_{0.20} s_{\bar{x}}$$

where:

CI = confidence interval

[8] The upper limit of the 80% confidence interval is equivalent to the one-sided 90% confidence interval on the mean, i.e., there is only a 10% probability that the true mean is greater than this limit.

DRAFT

$t_{0.20}$ = Student's t for the 80th percentile interval, obtained from Table 2 in Chapter 9 of SW-846 ($t_{0.20}$=1.638 for n=4, 1.363 for n=12, 1.300 for n=52, 1.282 for n>120).

$s_{\bar{x}}$ = standard error, as calculated by Equation 5 in SW-846, shown below

Equation (5), SW-846

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Because $t_{0.20}$ and $s_{\bar{x}}$ are both positive numbers, the upper limit of the confidence interval derived from Equation 6 above is:

Equation (B)

$$CU = \bar{x} + t_{0.20} s_{\bar{x}}$$

where:

CU = upper limit of the confidence interval

When the upper confidence limit is calculated for log-transformed data, the limit is translated back to arithmetic scale by the following equation:

Equation (C)

$$CU_y = e^{CU}$$

where:

CU_y = upper confidence limit on geometric mean (the median, the measure of central tendency for the lognormal distribution) translated back to arithmetic scale

Contaminants Monitored for TC Characteristic

This monitoring program is for benzene, the one TC constituent that may potentially exceed the TC limits in the Gallup Refinery wastewater discharge to AL-1. The remaining TC chemicals are not included in Western's routine monitoring program because they are either not expected to be present or may be present at concentrations well below their TC limits.

TC Calculations and Monitoring Frequency

TC monitoring is performed on the discharge from the air stripper system. A grab sample of the effluent from the air strippers is collected up to seven days/week when flow is present for the interim measures phase sampling. The benzene

DRAFT

concentration in each grab sample collected pursuant to this plan will be analyzed using an acceptable SW-846 analytical method.

The TC monitoring plan is intended to provide a methodology that allows continual observation of the performance of the air stripper system so that corrective action can be taken as necessary before the TC threshold is exceeded. It is similar to and is intended to be used as a quality control chart for a manufacturing process. Operations staff, with assistance from the individual(s) charged with maintaining the database, will be responsible for identifying increasing trends and taking corrective actions.

Recordkeeping

Documentation of analytical data used to classify a waste must include, at a minimum:

- Dates samples were collected,
- A description of the site or unit from which the sample is taken and sampling locations at the site or unit,
- Sample methods and sample equipment utilized, and
- Description of sample handling techniques, including containers, preservations, and chain of custody.

In addition, the records of TC calculations, including the rolling average, will be maintained.

The analytical data, including quality control data, will be maintained onsite and for at least 3 years after the waste is no longer generated or stored or until site closure. In addition, the following information must also be maintained for the same time period:

- Description of waste;
- Date of initial waste generation;
- Description of process that generated the waste;
- Hazardous waste determination; and
- Waste classification determination.

Historic Interim Measure Phase Results

Data have been collected during the interim measures phase beginning on January 10, 2010. A total of 425 samples were collected between January 10, 2010 and June 30, 2011 and analysed by Hall Laboratories in Albuquerque for benzene.

DRAFT

Figure 2 is a probability plot of the 425 samples collected between January 2010 and June 2011. The samples with measurable benzene concentrations (>0.01 mg/L) fit a lognormal distribution (see Figure 2). Approximately 23 percent of the samples were reported as having benzene concentrations that are below the reporting limit of 0.01 mg/L. Therefore, as described in this work plan, Cohen's MLE method was used with a lognormal distribution assumption to calculate the statistics of this benzene concentration data.

The calculated geometric mean of the benzene concentration distribution in Figure 2 is 0.045 mg/L. The 80th percentile UCL on the geometric mean is 0.048 mg/L.

DRAFT

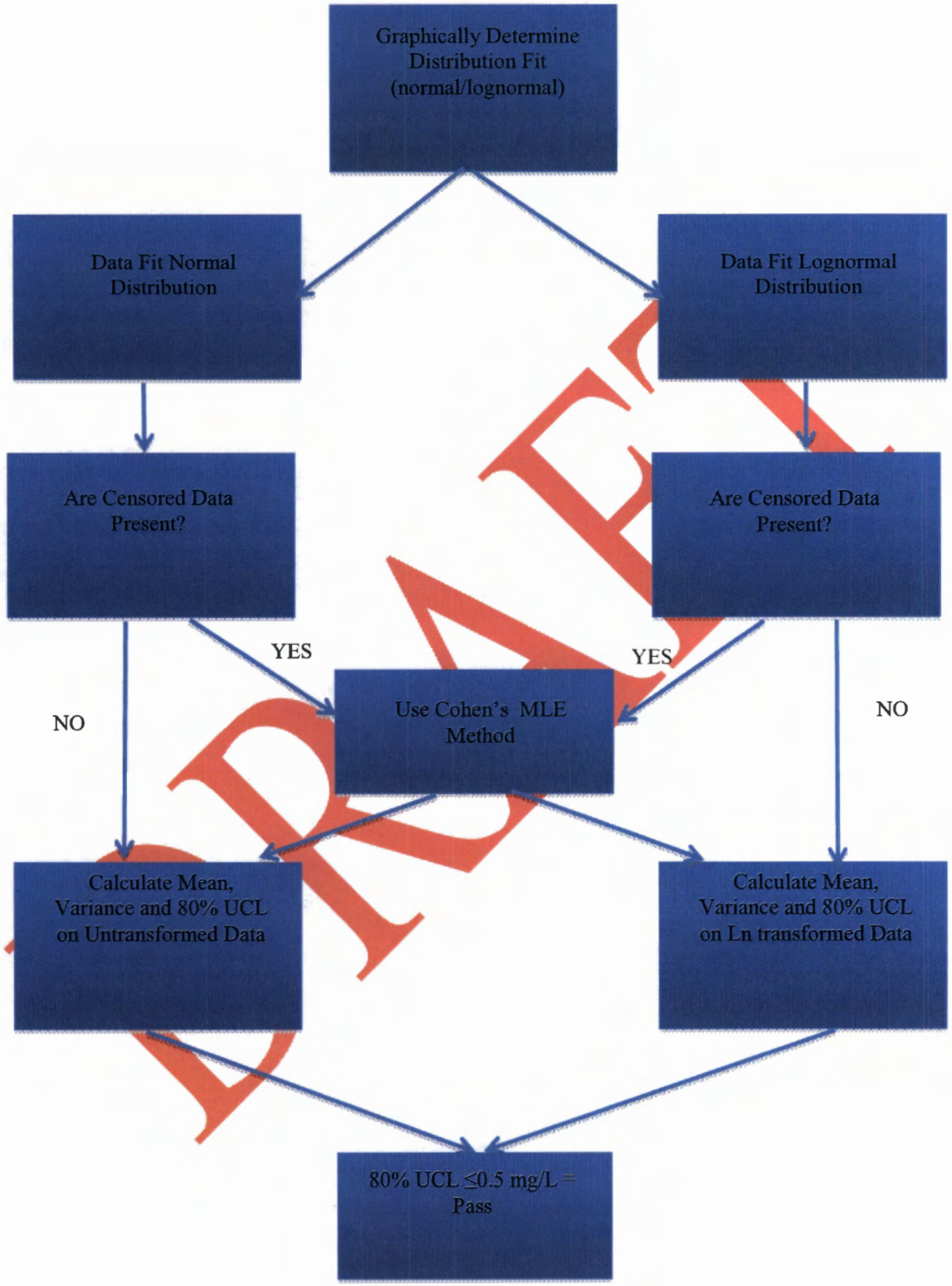
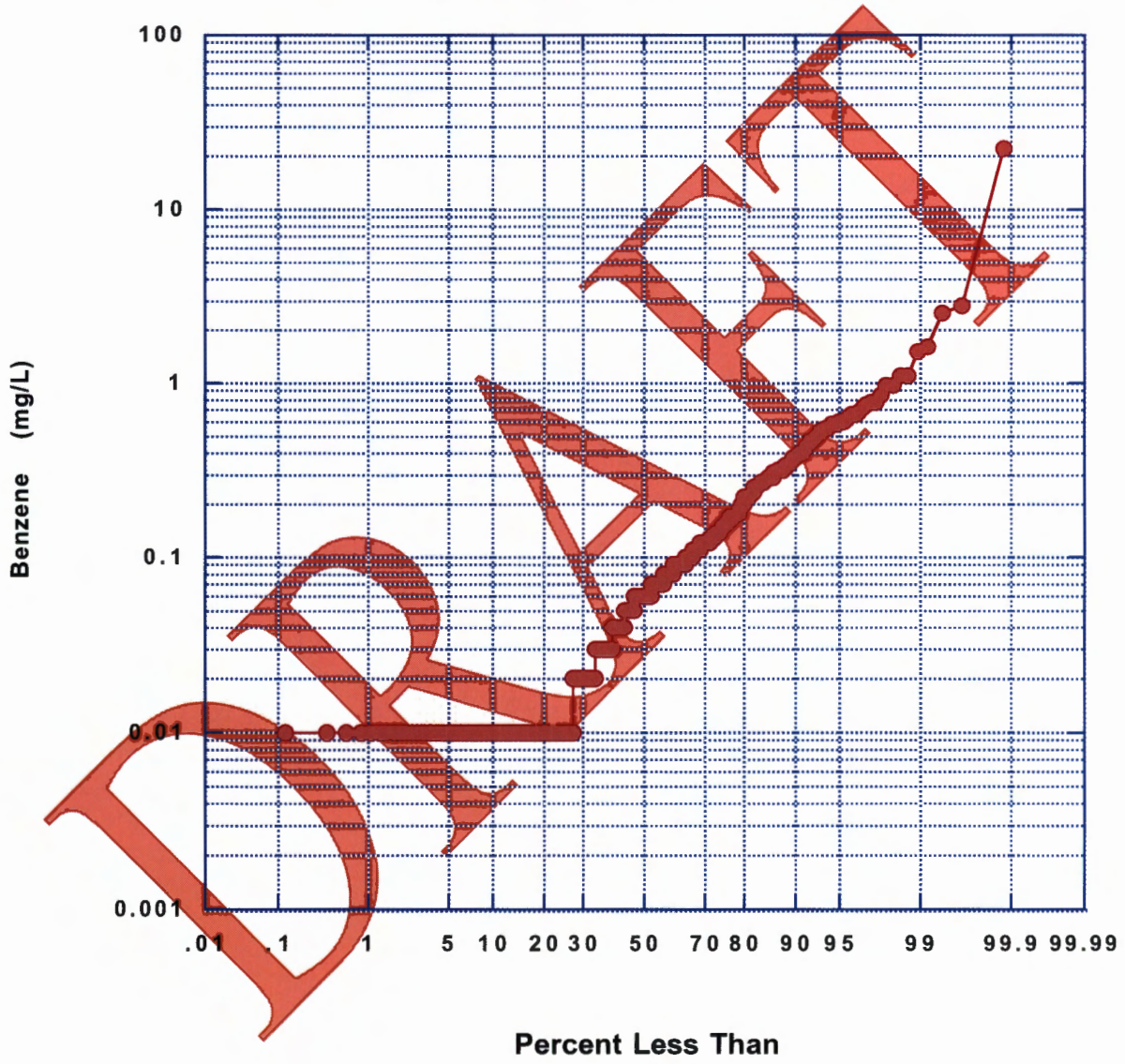


Figure 1
Flow Sheet for TC Statistical Data Analysis and Compliance Determination

Figure 2
Western Refining
Air Stripper Effluent Data Jan 2010-June 2011



DRAFT

Toxicity Characteristic Monitoring Plan Upgraded Wastewater Treatment System

Prepared for

Western Refining Southwest, Inc.
Gallup, New Mexico

by

Tischler/Kocurek
Registration No. F-3277
Round Rock, Texas

717
Version 1.0

July 2011

DRAFT

Western Refining Southwest Gallup Refinery
Upgraded Wastewater Treatment System
TC Characteristic Monitoring Plan

Introduction

This monitoring plan describes the procedures that are followed by the Western Refining Southwest, Inc. (Western) Gallup Refinery to monitor influent process wastewater from the Macro Porous Polymer Extraction System (MPPE) to the Sewer Treatment Pond-1 (STP-1) for evaluation of the Toxicity Characteristic (TC) for hazardous wastes.[1] Monitoring of the wastewater in accordance with this plan provides documentation that the plant wastewater (process wastewater and storm water from process areas), prior to entering the surface impoundment, is non-hazardous with respect to the TC.

This plan describes the toxicity characteristic, the regulatory basis for the monitoring approach, and how the monitoring data will be analysed to demonstrate continuing compliance.

The Toxicity Characteristic

RCRA Definition

The Toxicity Characteristic is one criterion by which a waste is defined as hazardous under the Resource Conservation and Recovery Act (RCRA). TC limits have been established for 40 compounds as shown in Table 1 at 40 CFR 261.24(b). If a representative sample of a waste equals or exceeds the TC limit, it is considered a characteristic hazardous waste and is subject to the applicable sections of the hazardous waste regulations.

Although the hazardous waste regulations do not specify the time period over which a representative sample must be calculated for a continuous discharge such as plant wastewater, TC limits were developed by the United States Environmental Protection Agency (EPA) based on a long-term exposure to contaminated ground water. The ground water model used by EPA assumes a chemical that leaches into the ground water is supplied by a constant mass flux from an infinite landfill or impoundment source (55 Fed. Reg. 11817) and that an individual's exposure over a lifetime is constant (55 Fed. Reg. 11823). Therefore, EPA assumed that an individual would be subject to continuous chronic exposure to a chemical by drinking contaminated water over the individual's lifetime (55 Fed. Reg. 11813). Accordingly, EPA developed long-term exposure limits for the TC chemicals and promulgated these in Table 1 of Part 261.

[1] 40 CFR (Code of Federal Regulations) 261.24

DRAFT

Representative Sample

In order to determine if a waste is TC hazardous, a representative sample of the waste must be compared to the TC limit (40 CFR 261.24(a)). A representative sample is defined at 40 CFR 260.10 as "a sample of a universe or whole (e.g., waste pile, lagoon, ground water) which can be expected to exhibit the average properties of the universe or whole." Regulations issued by the New Mexico Environmental Department (NMED) (20.4.1.100, 20.4.1.200 NMAC) incorporate the EPA regulations cited above by reference. Chapter 9 of EPA's *Test Methods for Evaluating Solid Waste* (SW-846) is the Agency's guidance for collecting representative samples.

EPA acknowledges in this definition that average[2] waste concentrations and not individual grab or composite samples must be used when determining if a waste is TC hazardous. This position is emphasized in Chapter 9 of SW-846, which states that: "it is not the variation of individual sample concentrations that is of ultimate concern, but rather the variation that characterizes $\bar{x}$ itself" ($\bar{x}$ being the average).

Other EPA documents provide additional regulatory guidance on representative sampling for hazardous waste characterization. In these, EPA points out the need to collect a representative sample, which may be developed from multiple grab samples, and using statistics to incorporate a level of confidence in the average.[3]

Chapter 9 of SW-846 provides guidance on the number of waste samples that are appropriate for obtaining a representative sample (Table 9-1) when data are available (i.e., when testing is used to determine if a waste is hazardous) to characterize the probability distribution of the waste characteristic to be measured. Equation 8 in Table 9-1 is:

$$n = \frac{t_{0.20}^2}{\Delta^2} \times s^2$$

where: n = appropriate number of samples
s = standard deviation of sample measurements

[2] The term "average properties" as used in the definition of a representative sample is the statistical measure of central tendency of the sample distribution that is representative of the population ("the universe or whole"). If the sample distribution of the data is normal, the arithmetic mean is the measure of central tendency. If the sample distribution of data is lognormal, the geometric mean (median) is the measure of central tendency. For distributions that are not normal or lognormal, the median or mode will be the measure of central tendency depending on the characteristics of the distribution.

[3] Letter from Mike Petruska, EPA Regulatory Development Branch, to Michael Roberts, Maintenance Engineering, June 22, 1995. EPA Memorandum, "Characterization of Nitrocellulose Filter Fabric Waste," from Elizabeth A. Cotsworth, EPA Acting Director, Office of Solid Waste to Barbara A. Finazzo, EPA Region 2 Director, Division of Environmental Science and Assessment, February 11, 1998. EPA, *Guidance for the Sampling and Analysis of Municipal Waste Combustion Ash for the Toxicity Characteristic*, EPA 530-R-95-036, Office of Solid Waste, Washington, D.C., June 1995.

DRAFT

$t_{0.20}$ = Student's t at 0.20 probability
$\Delta = R - \bar{x}$
$\bar{x}$ = mean of sample measurements
R = 0.5 mg/L (TC threshold for benzene)

No measured data will be available until the new system is started. Therefore, the initial number of individual grab samples that will used to calculate the statistics for the TC evaluation will be based on the effluent data from the interim phase air stripper system that was collected between January 2010 and June 30, 2011. A total of 425 samples was used for this calculation:

$\bar{x}$ = 0.20 mg/L
s = 1.0936 mg/L
$t_{0.20}$ = 0.8416 (424 degrees of freedom)
$\Delta = 0.5 - 0.2 = 0.30$ mg/L
n = 9 samples

The approval of the process design by NMED requires once/week sampling for benzene and selected other analytes by an independent laboratory.[4] Therefore, a 9-sample (9-week) moving average will initially be used to define a representative sample of treated effluent from the MPPE to determine if the discharge to STP-1 is below the benzene threshold for TC.[5]

Statistical Calculations

This section describes the statistical calculations for the rolling 9-sample average and how this average will be used to determine compliance with the TC concentration. Figure 1 is a simplified flow sheet of the data analysis process.

Fitting the Data to a Distribution

The first step in evaluating any data set is to determine what type of data distribution best fits the data because the type of distribution will affect the choice of statistics. With most data, the goal is to use normal distribution statistics. This is because there are many statistical tests for normal distributions and the calculations are relatively simple.

DRAFT

[4] Kieling, J.E., NMED, dated May 24, 2010, to Ed Riege, Western Refining Southwest, Inc., Re: Approval With Modifications Process Design Report for the Wastewater Treatment Plant Work Plan (Alternative Design, Revision A)

[5] Note: Operation of the new system will include twice/day sampling of the MPPE effluent and analyzed in the Gallup Refinery laboratory. If effluent from either the DGF or MPPE system is found to exceed 0.4 mg/L, the flows will be diverted. Wastewater transfer from T-35 and T-27/28 to the API separator will be shut off at the same time. This operational data will not be used to evaluate a representative sample of the discharge to STP-1, at least initially in the program.

If the original data do not fit a normal distribution, the data can be transformed by any number of functions to produce a normal data set. The most common transformation function for environmental data is the natural-log. So often do environmental data follow the lognormal distribution, that EPA recommends the distribution for daily values unless the distribution clearly does not fit the data (*Technical Support Document for Water Quality-Based Toxics Control,* (TSD), Appendix E). Two other data transformation methods recommended by EPA in SW-846 are the square root (when the mean is approximately equal to the variance) and the arcsine (when the mean is less than the variance).

The simplest, and often most effective way to determine goodness-of-fit, is to graph the data as recommended by EPA in its TSD (Appendix E) and in SW-846 (Chapter 9). If the data are normally distributed, they should plot as a straight line on a probability graph. If the data fit a lognormal distribution then the logarithms of the concentrations will be a straight line on a probability graph.

If the graphical probability analysis indicates that the sample data deviate significantly from a straight line on a probability graph, then a statistical test will be used to evaluate the goodness-of-fit to determine the appropriate distribution to be used for the calculation of the sample mean and variance. The Shapiro-Wilk test will be used to evaluate the goodness-of-fit for the normal and lognormal probability distributions when the visual comparison shows a significant deviation from normality. In such cases, the distribution that has the best fit to the sample data, as determined by the Shapiro-Wilk Test, will be used to calculate the mean and variance of the database for the confidence interval calculation.

A normal distribution is assumed for the data generated by this sampling plan. This assumption will be checked as data are generated. If the normal distribution is found to be unsuitable, the lognormal distribution will tested as a substitute. If neither the normal or lognormal distributions fit the database, a better fitting distribution will be selected.

Managing Censored Data

Concentrations that are reported as below the quantitation level of the analytical method for benzene are called *censored* values in statistical analyses. The simplest method for dealing with censored samples is to substitute one-half the reported quantitation (or detection) limit for each of the censored values in the database. The substitution method assumes that the concentrations below the quantitation limit are normally distributed and that the median/mean is the best estimate of the average concentration of the censored values in a data set. If the substitution method is used for databases with a large number of censored observations it will result in biased estimates of the mean and variance of the population. When a database contains a significant fraction of censored samples, a more robust statistical method for dealing with censored data is required. There

DRAFT

are several such methods described in the scientific literature,[6] and a commonly used method is Cohen's maximum likelihood estimator (MLE) method. Cohen's MLE method has been shown by a number of studies to provide acceptable estimates of the mean and variance of databases with a significant amount of censoring.[7]

This waste analysis plan will use the following decision criteria for censored data:

- If the number of censored samples constitute ≤ 10% of the total number of samples (i.e., 2 samples in the 9-sample database), the substitution method will be used and the statistics will be calculated as described below.

- If the number of censored samples exceeds 10% of the total number of samples in the 9-sample database, then Cohen's MLE method will be used to estimate the mean and variance of the sample population. The resulting estimated sample mean and variance will be used as described below to calculate the confidence interval that is used to assess compliance.

Descriptive Statistics

Equations from SW-846, as shown below, will be used to calculate the mean, the variance, and the standard deviation of the analytical data. According to SW-846, the equations shown are for simple random sampling as well as systematic random sampling. The monitoring data for the Western waste characterization are characterized as systematic random sampling, which according to SW-846, is when the first sample is randomly selected and all succeeding samples are taken at fixed time intervals. Because the hydraulic retention time in the collection system and treatment system (API separator, equalization tank, air strippers) are short, essentially any sampling frequency greater than a few hours will achieve the requirement for random samples.

<u>Sample Mean (arithmetic average)</u>

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Equation (2a), SW-846

where

$\bar{x}$ = sample mean
x_i = individual data values, i=1…n[8]
n = number of data points

[6] Berthouex, P.M. and Brown, L.C., 1994, *Statistics for Environmental Engineers*, CRC Press Boca Raton.
[7] Ibid.
[8] These data values will be transformed if necessary to meet the normal probability distribution assumption.

DRAFT

When the mean is calculated from data that have been transformed by the natural log function, it must be translated back to arithmetic scale by the exponential function as shown below. When the database fits a lognormal distribution, the mean that is calculated is the geometric mean (median) of the sample database, which is the measure of central tendency of a lognormal distribution.[9]

Translation of Log-Mean to Arithmetic Scale

Equation (A) $$\bar{y} = e^{\bar{x}}$$

where:

$\bar{y}$ = sample geometric mean translated to arithmetic scale
e = the exponential function, approximately 2.718281828

Sample Variance

Equation (3a), SW-846 $$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}$$

where:

s^2 = variance

Sample Standard Deviation

Equation (4), SW-846 $$s = \sqrt{s^2}$$

where:

s = standard deviation

Upper Confidence Limit

In Chapter 9 of SW-846, EPA suggests a method for adding a degree of conservatism to any calculated average when determining compliance with TC limits. EPA suggests that the upper limit of an 80% confidence interval[10] be

[9] Op cit., Berthouex, P.M. and Brown, L.C.

[10] The upper limit of the 80% confidence interval is equivalent to the one-sided 90% confidence interval on the mean, i.e., there is only a 10% probability that the true mean is greater than this limit.

DRAFT

calculated for the average and then compared with the TC limit. The equation in SW-846 for the 80% confidence interval is given below:

Equation (6), SW-846 $$CI = \bar{x} \pm t_{0.20} s_{\bar{x}}$$

where:

CI = confidence interval

$t_{0.20}$ = Student's *t* for the 80th percentile interval, obtained from Table 2 in Chapter 9 of SW-846 (e.g., $t_{0.20}$=1.638 for n=4, 1.363 for n=12, 1.300 for n=52, 1.282 for n>120)

$s_{\bar{x}}$ = standard error, as calculated by Equation 5 in SW-846, shown below

Equation (5), SW-846 $$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Because $t_{0.20}$ and $s_{\bar{x}}$ are both positive numbers, the upper limit of the confidence interval derived from Equation 6 above is:

Equation (B) $$CU = \bar{x} + t_{0.20} s_{\bar{x}}$$

where:

CU = upper limit of the confidence interval

When the upper confidence limit is calculated for log-transformed data, the limit is translated back to arithmetic scale by the following equation:

Equation (C) $$CU_y = e^{CU}$$

where:

CU_y = upper confidence limit on geometric mean (the median, the measure of central tendency for the lognormal distribution) translated back to arithmetic scale

DRAFT

Contaminants Monitored for TC Characteristic

This monitoring program is for benzene, the one TC constituent that may potentially exceed the TC limits in the Gallup Refinery wastewater. Samples are collected from the MPPE effluent before discharge to STP-1. The remaining TC chemicals are not included in Western's routine monitoring program because they are either not expected to be present or may be present at concentrations well below their TC limits.

TC Calculations and Monitoring Frequency

TC monitoring is performed on the discharge from the MPPE unit to STP-1. A grab sample (in duplicate) of the effluent from the MPPE is collected one day/week for analysis at an independent laboratory, as specified in the NMED May 24, 2010 approval letter. The benzene concentration in each grab sample collected pursuant to this plan will be analyzed using an acceptable SW-846 analytical method.

After one year's operation of the new treatment system the frequency of monitoring will be re-evaluated. The following guidelines will be used to establish the monitoring frequency based on the measured data.

		Maximum Concentration Measured in Past Year		
		<50% TC Limit	50-80% TC Limit	>80% TC Limit
Maximum Concentration of Any Sample in Past Year	Not detected	No TC calculations. Monitor once/month, or alternatively, maintain current monitoring frequency.		
	Detected, but all values <50% TC Limit	Calculate CU. Monitor at least twice/month.		
	>50% TC Limit	Calculate CU. Monitor weekly.	Calculate CU. Monitor 3/week.	Calculate CU. Monitor daily.

The TC monitoring plan is intended to provide a methodology that allows continual observation of the performance of the MPPE effluent so that corrective action can be taken as necessary before the TC threshold is exceeded. It is similar to and is intended to be used as a quality control chart for a manufacturing process. Operations staff, with assistance from the individual(s) charged with maintaining the database, will be responsible for identifying increasing trends and taking corrective actions.

DRAFT

Recordkeeping

Documentation of analytical data used to classify a waste will include, at a minimum:

- Dates samples were collected,
- A description of the site or unit from which the sample is taken and sampling locations at the site or unit,
- Sample methods and sample equipment utilized, and
- Description of sample handling techniques, including containers, preservations, and chain of custody.

In addition, the records of TC calculations, including the rolling average, will be maintained.

The analytical data, including quality control data, will be maintained onsite and for at least 3 years after the waste is no longer generated or stored or until site closure. In addition, the following information will also be maintained for the same time period:

- Description of waste;
- Date of initial waste generation;
- Description of process that generated the waste;
- Hazardous waste determination; and
- Waste classification determination.

DRAFT

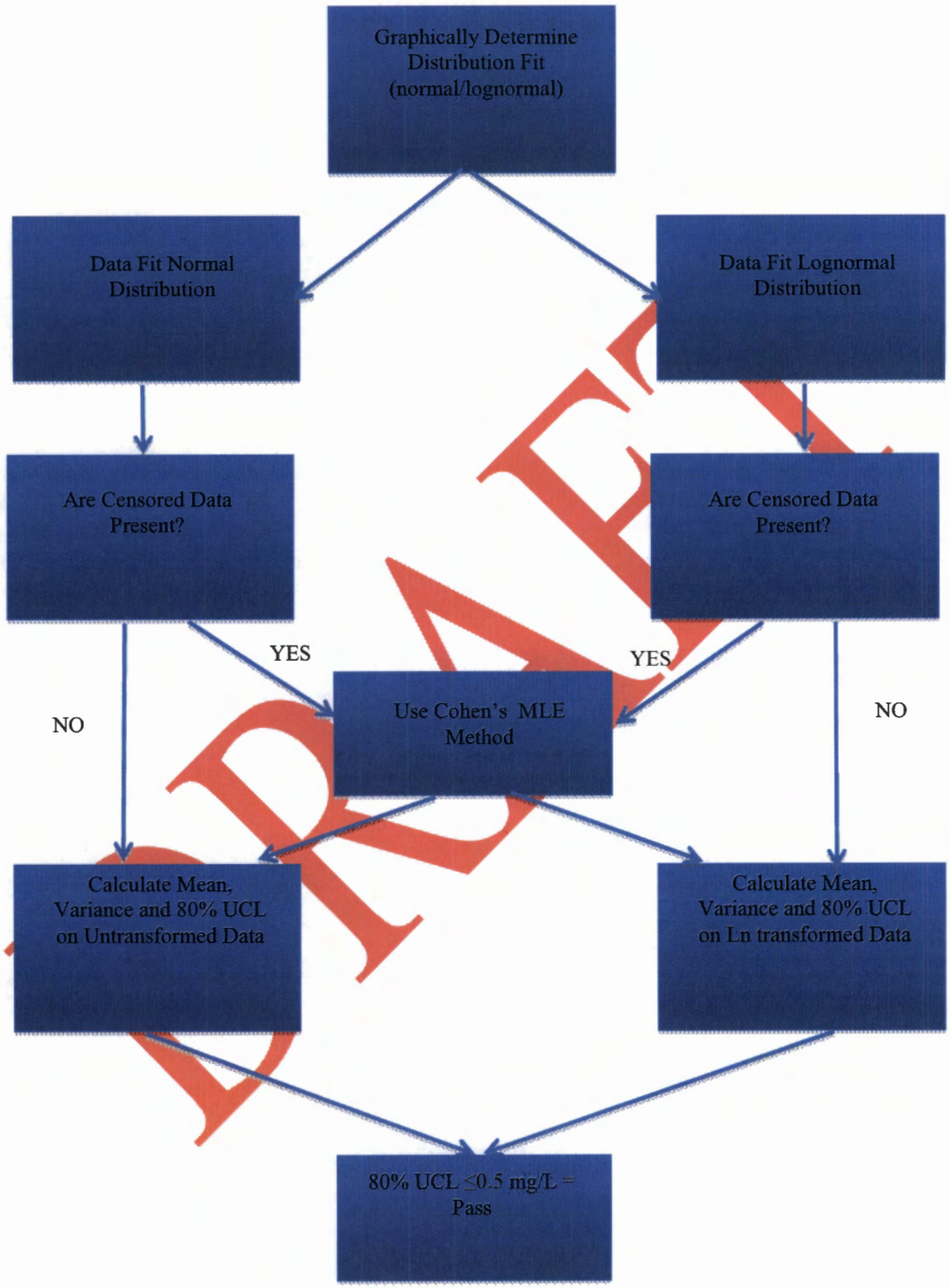


Figure 1
Flow Sheet for TC Statistical Data Analysis and Compliance Determination