UF | UNIVERSITY of FLORIDA

Center for Environment & Human Toxicology

PO Box 110885
Gainesville, FL 32611-0885
352-392-2243 Tel
352-392-4707 Fax

September 18, 2018

Leah J. Smith
District and Business Support Program
Division of Waste Management
Florida Department of Environmental Protection
2600 Blair Stone Road
Tallahassee, FL 32399-2400

Re: Literature review comparing the use of one-half the reporting limit to ProUCL methods of estimating non-detects

Dear Ms. Smith:

At your request, we have performed a literature review comparing the use of one-half the reporting limit to other methods included in ProUCL for estimating non-detects. The reporting limit is the smallest concentration that can be reported by the laboratory. In USEPA terminology, the reporting limit is the method detection limit, although in other contexts the reporting limit may be defined as a value greater than the limit of detection and less than or equal to the limit of quantification. For the purposes of this literature review, the term reporting limit is used to mean the concentration below which results are reported as "non-detect" and left censoring of the dataset occurs.

Several methods have been proposed for the estimation summary statistics in censored datasets. Each of the proposed methodologies has strengths and limitations in estimating the mean and variance of a population. No single method consistently outperforms the others in all instances due to differences in the percentage of censored data, number of reporting limits, sample size, and distribution of the data. Four of the most commonly used methods are summarized below.

<u>Simple substitution</u>

The simple substitution method consists of replacing the censored value with some fraction of the reporting limit. The most common substitution values include the reporting limit, one-half the reporting limit, the reporting limit divided by the square root of two, and zero. This method is based on the hypothesis that data below the detection limit follow a uniform (detection limit, detection limit divided by two) or triangular (detection limit divided by the square root of two) distribution (Baccarelli et al., 2005). Substituting one-half the reporting limit for non-detect data is not likely to bias the mean; however, it will decrease estimates of the variance and upper percentiles. Alternatively, assuming the non-detected value is equal to the reporting limit tends to overestimate the mean. Because the assumption is that all non-detected values consist of one value (the reporting limit), this method of substitution also decreases the estimates of variance and upper percentiles (ITRC, 2013; Hewett and Ganser, 2007). Substitution of censored data by one value limits the distribution and decreases the variance. Lognormally distributed environmental data are particularly sensitive to the choice of substitution values since the variance on these data sets is usually larger (Leith et al., 2010).

Several studies have shown that simple substitution methods do not work as well as methods that impute values for non-detects (Helsel, 2010; Leith et al., 2010; Hewett and Ganser, 2007; Singh et al., 2006; Baccarelli et al., 2005; Croghan and Egeghy, 2003). For data sets with a large number of non-detects, the substitution method distorts the data and produces incorrect results. Singh et al. (2006) studied simple substitution methods and showed that these methods also do not perform well even when the non-detected observations are as low as 5-10%. They recommend against using the substitution method for calculating 95% upper confidence limits (UCLs) on the mean in datasets with non-detected observations. This agrees with simulation studies performed by Helsel (2010) which showed substitution methods generally performed poorly. Based on Monte Carlo simulation results on PCB and p,p'-DDE concentrations in plasma collected from nestling bald eagles, Leith et al. (2010) found the one-half the detection limit substitution method produced a significant substitution effect for all datasets with greater than 11% non-detects. They conclude this is not a good method for analyzing left-censored data sets as the percent of non-detects increase. Croghan and Egeghy (2003) also used simulations to test the performance of simple substitution techniques. They determined that simple substitution is adequate for datasets with a slightly larger percentage of non-detects than the other studies. They concluded the error rate and relative difference in the means increase when the non-detected data exceed 25%. Their simulation also showed that replacement with the limit of detection divided by the square root of two produced better estimates of the mean than one-half the detection limit.

Strengths:

1. It is an easy methodology that requires little statistical knowledge.

Limitations:

1. Substitution introduces an arbitrary pattern into the dataset. If there are enough non-detects, this pattern introduces bias that can change the distribution of the data. This bias increases as the number of non-detects increase (Helsel, 2010; Leith et al., 2010; Helsel, 2005).
2. Because substitution limits the spread of the data below the reporting limit, it reduces the standard deviation and upper percentile values. Underestimating the upper percentiles in a dataset will also underestimate risk from exposure.
3. This methodology can result in inaccurate hypothesis tests (Helsel, 2006).

Kaplan-Meier Method

The most commonly used method to calculate summary statistics on censored datasets is the Kaplan-Meier method. The Kaplan-Meier method is a non-parametric method that relies on the ranks of the data, not any specific distribution. It uses this data to create a probability distribution function to impute the non-detects (ITRC, 2013). In an evaluation of several statistical treatments of left-censored data sets (including substitution, Kaplan-Meier, regression order statistics, and the maximum likelihood), Antweiler and Taylor (2008) concluded that for datasets with less than 70% censoring, the best overall technique for estimating summary statistics is the Kaplan-Meier method. Singh et al. (2006) performed an evaluation to compute the 95% UCL on datasets with multiple detection limits. Monte Carlo simulation experiments were performed on left-censored data sets over a wide range of skewed distributions. They found that the Kaplan-Meier estimate of the mean generally provided better coverage than the substitution, regression order statistics, and maximum likelihood estimates for data sets with up to 70% non-detects. The Monte Carlo simulations performed by Leith et al. (2010) showed Kaplan-Meier statistics provided the best estimate of geometric means in datasets with both left-hand censoring and right skew.

Strengths:

1. It is a non-parametric method so it does not require an assumption regarding the underlying distribution of the data. It is robust to all datasets where the true distribution departs significantly from a normal or lognormal distribution.
2. It can be used when multiple reporting limits are present.
3. Kaplan-Meier method can also be used when summing detect and non-detect data (such as in calculation of a dioxin toxic equivalence quotient (TEQ) or a benzo(a)pyrene TEQ (ITRC, 2013).
4. This method is not easily affected by outliers. Therefore, it is a good methodology for determining summary statistics on right-skewed datasets. (Leith et al., 2010; Antweiler and Taylor, 2008).

Limitations:

1. If only one reporting limit is present, the Kaplan-Meier method is equivalent to simple substitution because it will not estimate below the lowest detection limit. When performance is measured in absolute bias, the Kaplan-Meier method with one detection limit performs poorly when compared to regression and maximum likelihood estimate methodologies. Hewett and Ganser (2007) found this method resulted in a strong positive bias for the mean when only one limit of detection was present in the dataset.

To use the Kaplan-Meier method, a minimum of 8-10 data points with at least 3 detected concentrations are required. The non-detected concentrations should not make up more than 50-70% of the dataset (ITRC, 2013).

<u>Regression order statistics (ROS)</u>

Regression order statistics assume the data are normally, lognormally, or gamma distributed. This assumption is used to assign values to the non-detected concentrations in the dataset. Basically, this method calculates a linear regression line through the data (or log-transformed data for a lognormal distribution). The regression line is used to estimate the non-detected concentrations before calculating the summary statistics. This method assumes the non-detected data can be fit to a known distribution. It is considered a semi-parametric method because it only utilizes the distribution for the non-detected data and does not apply this assumption to the entire data set (ITRC, 2013). The ability of this method to estimate non-detects is dependent upon the fit of the regression line and the distance of individual data points from this line (Helsel, 2005). The normal, lognormal, and gamma ROS methods are available in ProUCL (USEPA, 2015a).

Strengths:

1. The gamma ROS method can be used with one as well as multiple reporting limits. This provides an advantage over the Kaplan-Meier method (USEPA, 2015a; ITRC, 2013).

Limitations:

1. Singh et al. (2006) noted that the ROS method does not provide adequate coverage of the mean for highly skewed left-censored distributions. This is true independent of the

size of the dataset and percentage of non-detected concentrations because the linear regression model produces a poor fit for highly skewed datasets.
2. Does not perform well in the presence of outliers. In the presence of outliers, estimates of the non-detects may become negative or exceed the detection limit (Singh et al., 2006).

To use the ROS method, a minimum of 8-10 data points with at least 3 detected concentrations are required. At least 50% of the dataset should consist of detected concentrations (ITRC, 2013). For more accurate and reliable results, a minimum of 10 detected observations should be used. Singh et al. (2006) recommend more than 10 detected observations when the percentage of non-detects is greater than 40%.

Maximum Likelihood Estimation (MLE)

Similar to the ROS method, the MLE approach also assumes the data are normally or lognormally distributed. This methodology uses an iterative process to solve for the mean and variance of the dataset (Croghan and Egeghy, 2003; Shumway et al., 2002). These statistics are calculated using parameters that represent the best fit to the distribution of the observed values (Antweiler and Taylor, 2008). Cohen's method is a frequently used form of MLE and assumes a normal distribution to the data. The Cohen's MLE method cannot accommodate more than a single reporting limit and can only assume the normal distribution. Because most environmental data are lognormal, the dataset needs to be transformed before estimating the mean and standard deviation. Transformation of the logarithms back into their original units introduces bias for datasets with less than 50 observations (Helsel, 2005). However, a study by Croghan and Egeghy (2003) concluded this method works better at estimating means than the simple substitution technique.

Strengths:
1. Performs well when the assumed distribution is correct. Helsel (2010) describes three different simulation studies conducted with colleagues in 1986 and 1988. These simulations showed the MLE method performed best for the estimation of percentiles as long as the assumption of the distribution was correct. If the assumption for the distribution was incorrect, the MLE method could produce percentiles that were very far off the true values.

Limitations:
1. If the wrong distribution is assumed for the dataset, the MLE method will not calculate accurate results. Only small departures from the assumed distribution are tolerated (Helsel, 2010).
2. Use of logarithmic transformations for environmental data can introduce bias, requiring large datasets (Helsel, 2005).
3. This method is sensitive to outliers (Antweiler and Taylor, 2008).
4. Does not accommodate multiple detection limits without decreasing performance (USEPA, 2015b).

The MLE method requires a large dataset. With multiple reporting limits, at least 50 data points and a detection frequency greater than 50% is required. Due to the limitations inherent in this approach, it is not included in ProUCL Version 5.0/5.1. However, it can be found in older versions of the program.

Please let us know if you have any questions regarding this review.

Sincerely,

Leah D. Stuchal, Ph.D. Stephen M. Roberts, Ph.D.

References:

Antweiler, RC, Taylor, HE (2008) Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets: I. Summary statistics. *Environ. Sci. Technol.* 42: 3732-3738.

Baccarelli, A, Pfeiffer, R, Consonni, D, Pesatori, AC, Bonzini, M, Patterson Jr., DG, Bertazzi, PA, Landi, MT (2005) Handling of dioxin measurement data in the presence of non-detectable values: Overview of available methods and their application in the Seveso chloracne study. *Chemosphere* 60:898-906.

Croghan, CW, Egeghy, PP (2003) *Methods for Dealing with Values Below the Limit of Detection using SAS.* North Carolina State University, Institute for Advanced Analytics, Raleigh, NC.

Helsel, DR (2005) More than obvious: Better methods for interpreting nondetect data. *Environmental Science & Technology* 39(20): 419A-423A.

Helsel, DR (2006) Fabricating data: How substituting values for nondetects can ruin results, and what can be done about it. *Chemosphere* 65: 2434-2439.

Helsel, DR (2010) Much ado about next to nothing: Incorporating nondetects in science. *Ann. Occup. Hyg.* 54(3): 257-262.

Hewett, P and Ganser, GH (2007) A comparison of several methods for analyzing censored data. *Ann. Occup. Hyg.* 51(7): 611-632.

ITRC (Interstate Technology & Regulatory Council). 2013. *Groundwater Statistics and Monitoring Compliance, Statistical Tools for the Project Life Cycle*. GSMC-1. Washington, D.C.: Interstate Technology & Regulatory Council, Groundwater Statistics and Monitoring Compliance Team. http://www.itrcweb.org/gsmc-1/

Leith, KF, Bowerman, WW, Wierda, MR, Best, DA, Grubb, TG, and Sikarske, JG (2010) A comparison of techniques for assessing central tendency in left-censored data using PCB and p,p'-DDE contaminant concentrations from Michigan's Bald Eagle Biosentinel Program. *Chemosphere* 80: 7-12.

Shumway, RH, Azari, RS, Kayhanian, M (2002) Statistical approaches to estimating mean water quality concentrations with detection limits. *Environ. Sci. Technol.* 36: 3345-3353.

Singh, A, Maichle, R, Lee, SE (2006) *On the Computation of a 95% Upper Confidence Limit of the Unknown Population Mean Based Upon Data Sets with Below Reporting limit Observations.* Lockheed Martin Environmental Services, Las Vegas, NV.

USEPA (2015a) *ProUCL Version 5.1 Technical Guide.* United States Environmental Protection Agency, Office of Research and Development, Washington, D.C.

USEPA (2015b) *ProUCL Version 5.1 User Guide.* United States Environmental Protection Agency, Office of Research and Development, Washington, D.C.