ATTACHMENT E

INTERIM APPROACH FOR EVALUATION OF UNCERTAINTY AROUND THE MEAN OF A SET OF ASBESTOS CONCENTRATION VALUES

1.0 INTRODUCTION

Human health risk assessments typically seek to estimate the risk to a person from exposure to a contaminant in an environmental medium at a particular location (often referred to as an exposure unit or exposure area). A key assumption in this approach is that exposure within the exposure unit is random (i.e., a receptor is equally likely to be exposed in all sub-locations within the exposure unit). Because of this assumption, the risk from a chemical in a medium is related to the arithmetic mean concentration of that chemical averaged over the entire exposure area. However, the true arithmetic mean concentration cannot be calculated with certainty from a limited number of measurements. Because of this, the USEPA recommends that the upper 95th percentile confidence limit (UCL) of the arithmetic mean at each exposure area be used when calculating exposure and risk at that location (USEPA 1992). If the 95% UCL exceeds the highest detected concentration, the highest detected value is used instead (USEPA 1989).

The mathematical approach that is most appropriate for computing the 95% UCL of a data set depends on a number of factors, including the number of samples in the data set, the type of distribution from which the samples are drawn, and the degree of censoring (occurrence of values below the detection limit). To help address these technical complexities, EPA has developed a specialized computer application referred to as ProUCL (USEPA 2007). This application calculates a number of alternatives estimates of the UCL for a data set provided by the user, and, based on the attributes of the data set, identifies which UCL estimate is most appropriate for use in risk assessment calculations.

An important assumption in the operation of the ProUCL software is that the concentration values provided in the data set are accurate. However, in the case of asbestos data sets, this assumption is not usually valid. This is because the concentration of asbestos in a sample is estimated by using microscopic techniques to determine the number of asbestos structures per unit medium. For example, for the analysis of an air sample:

$$C_{obs} = N_{obs} / V_{anal}$$

where:

C_{obs} = Observed concentration in air (s/cc)
N_{obs} = Number of asbestos structures counted during an analysis
V_{anal} = Volume of air (cc) that passed through the area of filter analyzed

However, the actual number of asbestos structures observed during an analysis is itself a random variable characterized by the Poisson distribution:

$$N_{observed} \sim \text{Poisson}(N_{expected})$$

where:

$$N_{expected} = C_{true} \cdot V_{anal}$$

Thus, the observed concentration is a Poisson random variable given by:

$$C_{obserevd} \sim \text{Poisson}(C_{true} \cdot V_{anal}) / V_{anal}$$

For example, assume that the true concentration of asbestos in a sample of air was 0.026 s/cc, and that 100 L of air was passed through a filter with an area of 100 mm^2. This would result in an average loading on the filter of 26.0 s/mm^2. If a total area of 0.1 mm^2 of filter (about 10 grid openings) were examined by the microscopist, it would be expected that the number of structures observed would be, on average across multiple analyses, about 2.6. However, in any one analysis, the number of structures observed can not be 2.6, but must be an integer (0, 1, 2, 3, 4, 5, etc.). The relative probability of observing each of these alternative counts is shown in Figure 1.

This Poisson measurement error in each sample concentration estimate, superimposed on the random sampling variability between different samples, results in a "mixed distribution" with complex statistical properties. For example, if the true distribution of air concentrations were characterized by a lognormal distribution, the sample set available for analysis would be a Poisson-lognormal (PLN).

Because ProUCL assumes that the only source of variation between samples is sampling error, and was never intended to deal with the case where measurement error was also occurring, use of ProUCL to estimate the 95% UCL of an asbestos data set does not yield results with the desired statistical properties (95 out of 100 UCL values exceed the true mean). Rather, the probability that the computed UCL exceeds the true mean is decreased, with the magnitude of the decrease depending on the size of the data set, the average number of counts, and the number of samples with a count of zero.

At present, the EPA is working to develop a computer application that is similar in concept to ProUCL, except that it is capable of dealing with mixed distributions that contain Poisson measurement error. However, this new application is not yet available for use. Therefore, an alternative approach must be used until that application is available. The purpose of this Attachment is to describe an interim approach that has been developed for use at the Libby Superfund site. This method is a screening approach that is not statistically rigorous, but which

does provide a semi-quantitative basis for estimating the magnitude of the uncertainty in the observed mean of an asbestos data set.

2.0 INTERIM APPROACH

One approach for characterizing the uncertainty around the mean of an asbestos data set is to use Monte Carlo simulation. In this approach, it is first necessary to specify the "true" distribution of concentration values in the medium being sampled, and to specify the amount of sample analyzed. Given this, a large number of alternative data sets may be generated, and the variability in the observed means may be characterized as a function of the average number of particles counted per sample, the number of samples in the data set, and the variability between samples in the data set.

In general, the true underlying distribution that characterizes the variability between different samples in a medium is not known. However, in most cases, the distribution is right skewed, and is often reasonably approximated by a lognormal distribution. If the true distribution is less skewed than a lognormal, then assumption of lognormality will generally tend to overestimate uncertainty somewhat. Based on this, for this interim method, the underlying distribution of concentration values was assumed to be lognormal. The between-sample variability in a lognormal distribution is characterized by the geometric standard deviation (GSD). For the purposes of this effort, three alternative GSD values were evaluated (3, 6, and 10). A GSD of 3 is suspected to be at the low end of the range of values that might occur at Libby, while a GSD of 10 is suspected to be near the high end of the likely range. The number of samples in each data set was assumed to vary from a small number (5) up to a high number (80). Each sample was assumed to be analyzed by a procedure with random Poisson counting error, with the average number of particles counted per analysis ($\lambda = Ctrue \cdot V_{anal}$) ranging from 3 to 20. The mean of each simulated data set was computed, and divided by the true mean in order to normalize the values.

3.0 RESULTS

The results of the simulation (presented as the range from the 5th percentile to the 95th percentile of the ratio of the simulated mean divided by the true mean) are shown in Figure 2. As expected, the width of the sampling distribution of the mean tends to decrease as sample number increases, and to increase as a function of underlying variability (GSD). Based on this figure, the approximate uncertainty around the observed mean of an asbestos data set may be assessed as follows:

1. Compute the observed mean and GSD of the data set.
2. Based on the GSD_{obs}, estimate the potential magnitude of the uncertainty in the mean based on the sample size and the GSD_{obs}. For example, for a data set of size 20 and a

GSD_{obs} = 6, it may be estimated that the true mean is probably within a factor of about 3 (either higher or lower) of the observed mean.

3. Using the high end of the uncertainty range, estimate a "high end" value for the mean. In the example above, the high end estimate would be 3-times the sample mean.

As emphasized above, this simulation approach does not constitute a rigorous method for computing the 95% UCL of a given data set. The chief limitation to the method is that it assumes the observed sample GSD is reasonably close to the truth. However, the value of GSD_{obs} is a random variable, and the true GSD might be either higher or lower than the value of GSD_{obs}. Consequently, the high end estimate for a data set might be either higher or lower than a rigorously computed UCL for that data set.

REFERENCES

USEPA. 2007. ProUCL Version 4.00.02 User Guide. U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory. EPA/600/R-07/038. April 2007. Available online at www.epa.gov.

FIGURE 1
EXAMPLE POISSON DISTRIBUTION

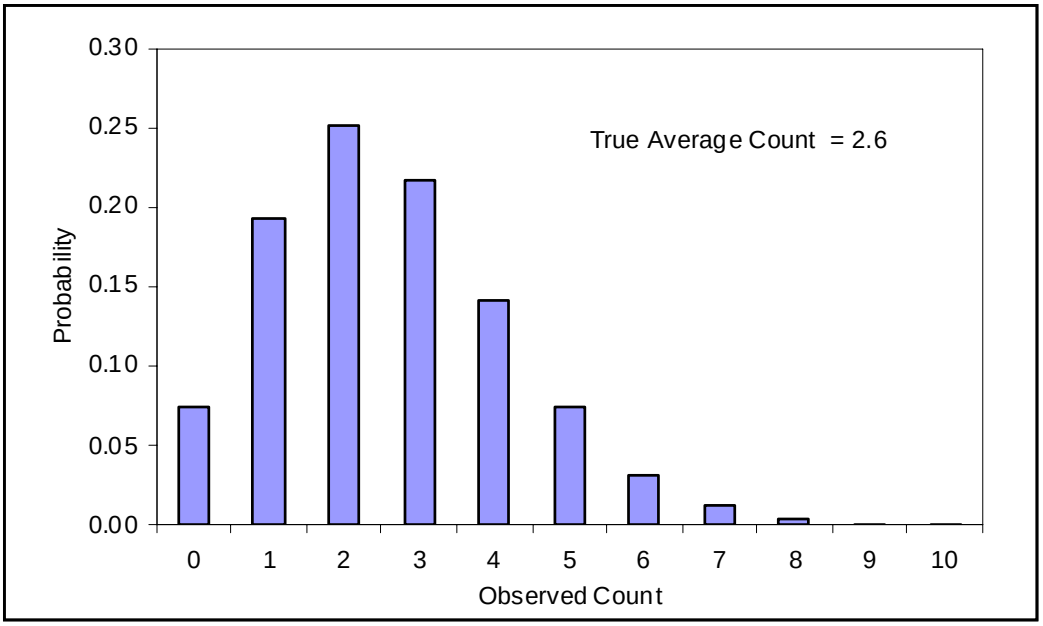
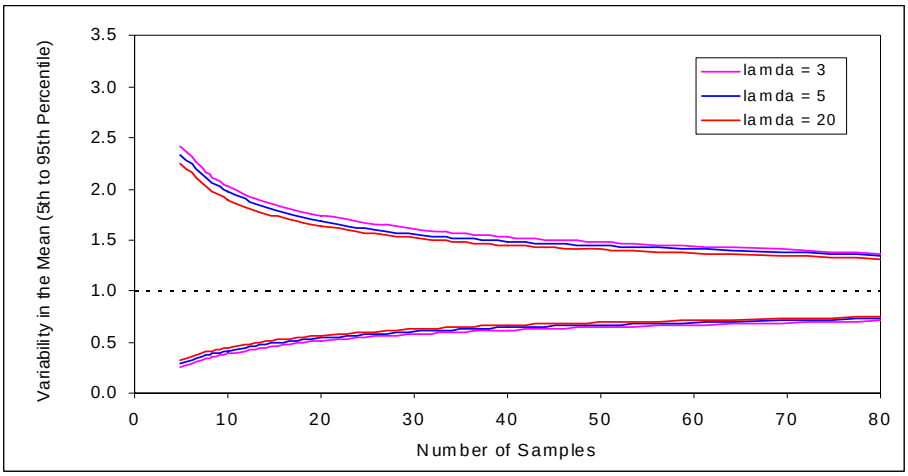
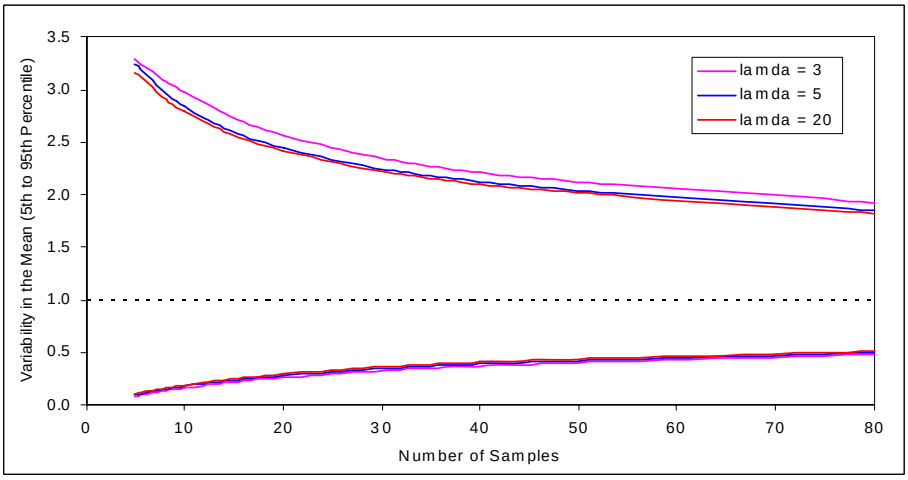
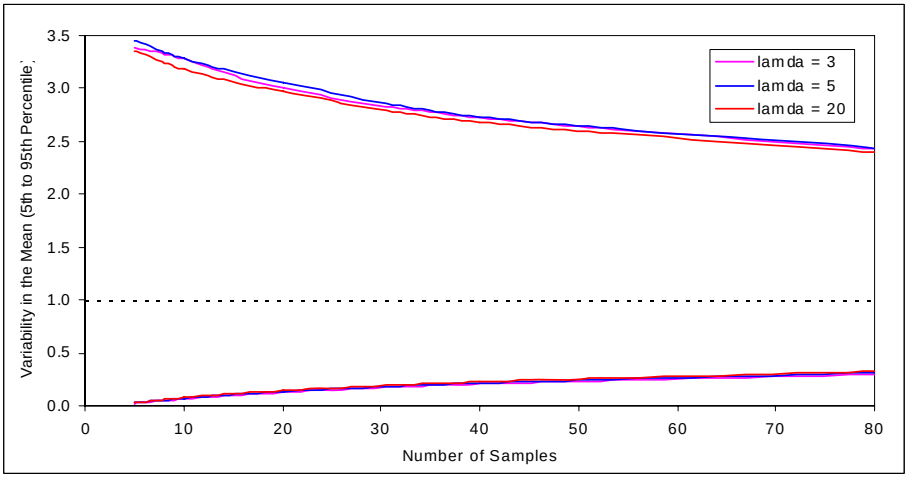


FIGURE 2
SIMULATED UNCERTAINTY IN THE MEAN OF A PLN DATA SET

GSD = 3



GSD = 6



GSD = 10