C10 On Confidence Intervals Used to Determine Background Concentrations

Willem A. Schreuder, PhD, Principia Mathematica Inc., 445 Union Boulevard, Suite 230, Lakewood, CO 80228*

The goal of this presentation is to make attendees aware of the consequences of including and excluding outliers including non-detects when determining background concentrations.

This presentation will impact the forensic science community by showing what confidence intervals represent.

Determination of a background concentration is often a controversial exercise. It is further complicated by noisy data which requires a statistical expression of what the background concentration measurements mean, typically expressed as using a confidence interval.

A confidence interval simply expresses to probability that the mean value of the measurements will be less than a given value. When a finite set of measurements of the same quantity (background concentration in this instance) is made, the mean value represents the most likely (expected) value of that quantity. It is possible, however, that the mean value of a finite set of measurements does not represent the true mean of the entire population, in this instance all the possible samples that could be taken. To express the certainty with which we can say that the set of samples we do have represents the true value you would find with a much larger (or infinite) set of samples, it is customary to express the result in terms of a confidence interval.

When the measurements are consistent, that is, the standard deviation of the measurements are small, the confidence intervals are narrow. Specifically, if all the measurements are exactly the same, the standard deviation is zero, and it is extrapolated that no matter how many more samples are taken, the measured value would always remain the same. However, when the data are noisy and have a large standard deviation, the confidence intervals can be very large. For such data, the actual population of concentrations vary and exactly which subset of all the possible subsets we measure could give very different answers for the mean. As such, the large confidence interval expresses the fact that the true mean of the entire population may be poorly captured by the actual subset of measurements taken.

Special care should be taken with the treatment of extreme values, sometimes called outliers, as such values would increase the standard deviation of the data and hence the confidence intervals. On the low end of the concentration range are often non-detected values, while at the high end there may be outliers which may actually be the result of sample contamination or influence by the source being investigated.

Where background concentrations are relatively low, it is common to find a number of non-detects in the data set. Since concentrations are strictly positive and often show a log-normal distribution, selecting the specific concentration to use for the non-detect is very important. Half the detection limit is often used; however, when the detection limit varies among the samples, this may introduce an artificial variation in the sample set since the detection limit simply expresses an upper bound on the true concentration, while the actual concentration may show little variation among samples. Conversely, it is possible that the actual concentration varies greatly among the samples, but that this concentration is below the detection limit and that the same detection limit is reported for many samples. This would artificially reduce the variability.

This presentation discusses the statistical basis of confidence intervals, and demonstrates using examples the consequences of different ways used to treat these outliers.

Confidence Interval, Background Concentrations, Statistics

 ** Presenting Author*