A105 Statistical Methods for Reducing Inaccurate Bias During Manipulation of Data Below the Detection Limits in Forensic Investigation

Robert Posey, MChem, University of East Anglia, Earlham Road, Norwich, NR4 7TJ, UNITED KINGDOM; and Jurian A. Hoogewerff, PhD, Centre for Forensic Provenanging, School of Chemical Sciences, University of East Anglia, Norwich, Norfolk NR4 7UA, UNITED KINGDOM*

After attending this presentation, attendees will be introduced to the risk of data set biasing by incorrect treatment of results below the experimental detection limits (censored data) and will leave with knowledge of a number of data handling techniques suggested in a number of environmental sources for the treatment of such data. Attendees will also be introduced to some powerful statistical software to facilitate the use of these techniques.

This presentation will impact the forensic community to the need for more universal methods for the analysis of non-detect data in order to prevent ambiguity between groups, and to reduce the risk of biasing forensic data sets which could lead to inaccurate conclusions. It will introduce techniques available for the analysis of non - detect data that can result in more accurate conclusions that the commonly used detection limit substitution methods.

Forensic investigation often involves the analysis of materials containing only trace levels of the analyte of interest. Data sets from such investigations often contain values that lie below the detection limit of the analytical method. Traditional methods for dealing with these values involve either substitution of a value as a function of the detection limit, (essentially fabricating the data based on no theoretical data), or by removing the value altogether. These techniques can significantly bias the summary data for a data set due to the population distribution becoming skewed. A number of statistical techniques have been proposed by the environmental community for the calculation of summary data for data sets containing censored data values based on the population distribution of values within the data set above the detection limits. There are also methods proposed for the calculation of summary data for data sets containing only censored data. Of particular interest are the methods described by D.R. Helsel (2005) which can easily be performed using the NADA for R macro in the powerful statistical software R.

The author will present results for the concentrations of trace elements in a selection of 104 European olive oils from 5 countries, which were measured using a novel LA-ICP-MS technique developed at UWA funded by the European Commission TRACE project. Olive oils are notorious for containing very low levels of trace elements (ppb) and the novel laser ablation technique was proposed as a possible improvement on traditional sample dilution techniques. A number of groups have reported successful differentiation of olive oils from different regions but only on very small sample sets and a limited number of elements suggesting stochastic difference causing the differentiation rather then a deterministic process. Our sampling scheme should therefore represent a more realistic approach. As expected a high percentage of the data was censored. Summary data was obtained for 25 elements using non-detect analysis techniques. Between group significance testing was performed using Peto-Prentice, MLE and ANOVA methods and concluded that no differentiation could be made between the geographical regions using the data available. Analysis was made possible by the implementation of the non detect analysis techniques without which little conclusion could be made from a data set with such a high percentage of censored values.

Non-Detect, Statistics, Bias

 ** Presenting Author*