C15 QA/QC and Data Defensibility for Environmental Forensics

Jennifer L. Holmes, PhD, Severn Trent Laboratories, Inc.,5755 8th Street East, Tacoma, WA 98424; and Andrew John Friedman, BS, BA, OGW Research Laboratories, 3635 Woodland Park Avenue, North, Seattle, WA 98103*

Attendees will gain knowledge of the necessity of evaluating data quality for answering environmental forensic questions that may or may not have been part of the original data sampling and analyses objectives. This presentation will impact the forensic community by providing a better understanding of how complex it is to consider data obtained for one purpose toward another.

It is critically important when assessing data for forensic purposes to assure that the data are suitable for answering the questions involved in the case. When environmental data is generated it is commonly done by laboratories whose procedures are designed to satisfy regulatory requirements for state, federal and/or tribal agencies. The EPA, state and other standard methods often state the frequency and criteria for the QA/QC samples to establish a likely concentration range for analytes above the practical quan- titation limit. These data are often used subsequently for litigation purposes that need to answer questions that are different than those for which the analyses were designed.

The question of whether the methodology-generated QA/QC para- meters are sufficient for answering the litigation question(s) is an issue that needs to be carefully investigated before the analyses are used for these new purposes. All too often the person originally requesting and using the data does not take the QA/QC into account except to the extent that it did or did not pass some standard criteria designed to assure that it is suitable for the original purposes the analysis was intended to address. When a sub- sequent user, say a geologist or environmental engineer is evaluating the data, the concentrations as if true at all levels and covering a wide range of possibilities are often used. This is not always the case.

Consider the different QA/QC samples.

Blanks: Typically the field QA/QC samples are examined. The field blanks and the trip blanks are generally evaluated and, if contamination is shown, the samples are then examined as the collection or shipping may have led to contamination. However, often overlooked are the various lab- oratory blanks, such as prep blanks, instrument blanks and calibration blanks. These are not always as closely examined, unless the data goes to court. If the calibration blank has an absolute value (note it is an absolute value so it can be above or below zero) that is greater than the reporting limit then the sample result for low level analyses is likely to be biased. The biases can be either high or low, but the result of the sample cannot be considered to be a “true” value but rather a limit. Considering the value as a limit may restrict the relevance of the data itself.

Laboratory duplicates and matrix spikes*:* It is difficult to draw con- clusions regarding other samples based on the results of only one or two in every batch or group of twenty samples. If the QA/QC sample sent to the laboratory is a composite then it may be more scientifically logical to draw conclusions regarding all samples, but still it is difficult to argue that the individual samples behaved in the same way. Unfortunately, labs tend to treat QA/QC samples differently simply because labs are aware of their importance. This can show up in such differences as that the sample and duplicate are homogenized longer, or the matrix spike may be added after the sample is digested. Unless the lab fully documents the entire digestion and spiking process it is difficult to know and the spike is added prior to digestion, the matrix spike may be more carefully digested or treated. The logbooks that are used to record such data at times may lack information such as the amount of time the digestions was given or the individual bomb temperatures. Typically logbooks are used as a way to record sample ID and data such as the amount of mass used, however the logs are not a check list to ensure that all of the steps were completed. It is therefore important to review the written procedures that the lab actually used, rather than going to the generic source (the guidelines) to verify the adequacy of the procedure.

Certified reference materials (CRM): These samples (or laboratory blank spikes) often represent the absolute best that a lab can achieve. Considering that CRMs are either reagent water or freeze dried matrices and that laboratory blanks are always reagent water, this is not surprising. The lab may have criteria for CRMs and laboratory blanks that are similar to matrix spike recoveries (typically 75-125%) however those criteria are for real samples, which CRMs and reagent water are not. Reagent water blanks spikes should be able to achieve as good a result as a calibration standard. Typically, calibration standard results are held to 90-110% recovery. Considering that CRMs are not reagent blanks they may not have such good of recoveries, however they are very homogenous samples that are not really field samples, the recoveries should be considered the best achievable with the method. Like the spiked samples CRMs are often treated more carefully than field samples, including possibly being digested or extracted independently of the field samples being examined, making the actual recoveries of the CRMs less reliable than matrix spikes of the same material as indicators of how the analyses went.

Other Factors Affecting Analyte Quantitation:

Even within the regulatory guidelines, labs often use modified methods, providing the methods demonstrably deliver similar results for test samples used to audit laboratory performance. This means that the lab has modified the method, which may include the frequency and/or the cri- teria for the QA/QC results as well as any corrective actions to undertake and it is important to assess what effect these modifications will have under the

 ** Presenting Author*

requirements of the new questions, not just the generic specifications of the guidelines.

When laboratory samples cannot achieve a 90-110% recovery the sample results should be considered to be suspect in that the actual value could be different than the reported value. Trends should be considered as well. Low recoveries can mean that the method itself has bias. However it can also be an indicator that the lab is using a solution that may be near or past expiration especially in the case of organics. High recoveries can be an indicator that the laboratory is biased or that the laboratory itself has contamination issues.

Laboratory contamination is always a concern. Laboratories have the analytes of interest in high concentrations to make standards and spiking solutions so contamination is always a possibility. Often the sampling con- tainers are purchased from the laboratory as well. If the method by which the containers are cleaned as well as how "clean" the containers are is not documented then the samples must be considered suspect and any contam- ination could be argued to be due to the laboratory or sample containers. Thus once again the data is not as defensible.

Perhaps the most important matter is to examine any surrogates, internal standards or other standards to check performance. The question now being addressed may deal with different parameters than the original data objectives. It may be the checks, surrogates and/or internal standards are not relevant or suitable for the new question being examined. This will definitely pull into question the reliability of the data and must be con- sidered before drawing interpretative conclusions from this data. The data may still be useful or not, depending upon the actual questions now being considered.

The above checks need to be completed before the data are evaluated for nonrandom variations that are attributable to batch or operator varia- tions over the course of the study. These are often hidden variables because they are not documented in many data reports except in the run logs of the laboratory. Since samples are usually submitted in groups in a nonrandom fashion over the course of site investigations, any variations due to machine performance, extraction variations (including personnel), calculation stan- dards and algorithms have to be accounted for before any more advanced examination of statistically significant data differences can be done. When this is not the case, the sorting of data for statistically significant trends may actually be sorting out nonrandom biases in the data sets.

Examples of all of these problems will be presented during the course of the talk leading to the inevitable conclusion that it will always be a complex issue to determine if sample data is reliable and reasonable. All the QA/QC samples as well as historical data and sample data need to be examined for trends, bias, or any peculiarities. Once the reliability of the data has been assessed, and only then, can the relevance of the data toward the questions being asked be carefully addressed. Data acquired for moni- toring purposes may or may not answer questions regarding remediation, contamination extent or identify the potentially liable parties.

QA/QC, Evaluation, Significance

 ** Presenting Author*