

**APPLICABILITY OF LABORATORY DATA GENERATED FOR  
COMPLIANCE  
WITH SAFE DRINKING WATER REGULATIONS**

**WHITE PAPER**

Prepared for:

**National Rural Water Association  
2915 South 13<sup>th</sup> Street  
Duncan, Oklahoma 73533**

Submitted By:

Sanford Cohen & Associates, Inc.  
(SC&A, Inc.)  
1000 Monticello Court  
Montgomery, Alabama 36117

By:

Name: Charles R. Phillips and Donald A. Chambless

**May 8, 2007**

# Applicability of Laboratory Data Generated for Compliance With Safe Drinking Water Regulations

## EXECUTIVE SUMMARY

The purpose of the present study was to obtain and evaluate drinking water compliance data from a representative number of states to assess the total variability and reliability of these reported data. Specifically the results of this study provide a representative determination of the statistical validity of using the mean of four-quarterly measurements for determining compliance with drinking water regulations.

Under the process currently used to determine compliance of a water system with respect to a given analyte, four-quarterly measurements are averaged and their mean is compared to the relevant MCL. If the mean does not exceed the MCL the water system is found to be in compliance, but, otherwise, the system is found to be noncompliant and remediation may be required, often at great cost to the water system and, therefore, to its customers. In some instances where the mean exceeds the MCL, systems are allowed to resample to verify the mean value obtained from the four-quarterly measurements.

In this work, SC&A gathered a large amount of compliance data for many analytes and water systems. Since the matter of interest here is the stability and reliability of the mean of the quarterly measurements as a measure of the levels of various analytes in the water supplied by water systems, SC&A made comparisons, for each analyte and each water system, between the mean analyte level and its associated uncertainty,  $E$ , which SC&A determined using standard statistical methods.

Graphs comparing the uncertainty,  $E$ , to the mean,  $\bar{x}$ , were constructed for those water supply systems and those analytes where we were able to obtain at least two consecutive sets of four measurements. These graphs are contained in Appendix A. We believe that an elaborate quantitative analysis is not in order for this somewhat preliminary study and that a simple examination of the accompanying graphs will lead the reader to the appropriate conclusions presented below.

The results of this comparative study revealed that the statistical uncertainty in the mean often exceeds the mean itself, i.e., the uncertainty of the mean is frequently more than 100% of the mean. The importance of this is difficult to overstate. If one thinks of quantities subject to statistical uncertainty as being “fuzzy,” then the mean of the quarterly measurements is frequently so fuzzy that a distinction between the three concentrations (1) zero, (2) the mean, and (3) twice the mean cannot be confidently drawn. Clearly, to make a decision to require remedial water treatment action in the presence of such uncertainty is ill advised, if not illogical.

To assist the reader in visualizing the degree of uncertainty in four consecutive measurements, the following table was prepared, based on the data used to generate the graphs in Appendix A. The table presents the average uncertainty ratio ( $E/\bar{x}$ ), expressed as percentage, for each of the

analytes represented in the graphs and the number of sets of four measurements (repeated or quarterly) whose uncertainty ratios are included in the average.

	<b>Arsenic</b>	<b>Copper</b>	<b>Fluoride</b>	<b>Gross Alpha</b>	<b>Ra-226</b>	<b>Ra-228</b>	<b>TOTAL HALOACETIC ACIDS (HAA5)</b>
Mean Uncertainty Ratio (%)	46.6	33.8	48.2	82.8	72.7	73.6	54.3
Number of Data Sets	30	5	4	80	23	19	20

Just as in the graphical representations, the high degree of uncertainty is obvious when presented in tabular form, as above.

There were a significant number of instances in the data where multiple results were listed for a given water source and analyte with reporting dates separated by only one to two days. It is not known nor did we attempt to determine the reasons for these multiple “short-term”, also referred to as “repeated”, measurements. When we applied the same statistical analysis to these short-term (repeated) data as for the four-quarterly ones, there appeared to be no discernable difference in the variability (reliability) between it and the quarterly data. This finding brings into question the reliability of compliance data and its usage for enforcement purposes; it suggests that the sampling and analytical processes may be incapable of supporting enforcement.

The results of this study should be considered preliminary, and the subject deserves additional study. By using the available resources and investing a significant amount of time and effort into acquiring the data, we were able to obtain sufficient data to examine the fundamental statistical properties of compliance data. However, to assure that our preliminary findings are indeed representative, a much larger sample of data, extended to cover a greater geographical area and additional analytes, should be subjected to a more quantitative analysis.

## **1.0 Introduction**

The assessment of water quality is typically made by comparing the mean of four-quarterly measurements with the relevant maximum contaminant level (MCL). If the mean of the quarterly measurements for an analyte exceeds the associated MCL, the water system is judged to be out of compliance with regard to that analyte<sup>1</sup>, and the system may be required to implement additional treatment measures. The costs of procuring, installing, and maintaining the equipment necessary to perform such additional water treatment can be quite large, and this often constitutes a significant hardship on the smaller water systems—in the case of the larger systems

---

<sup>1</sup> In some instances where the mean exceeds the MCL, systems are allowed to resample to verify the mean value obtained from the four-quarterly measurements.

this cost is spread over a much larger number of consumers and thus has a much lower adverse impact on the individual customers of these larger systems. Since water consumption costs are largely nondiscretionary, the customers of the smaller water systems have little opportunity to reduce the impact of the costs of the additional treatment equipment; the costs of such equipment must, of course, be passed down to the customers by the water system.

While no ethical professional would seek to avoid implementing truly necessary remedial treatment procedures, the determination of the necessity for such additional treatment should be based on the use of data of known quality and with a clear understanding of the limitations of the testing procedures. The current procedures used for determining compliance do not take into consideration data quality in any manner whatsoever—it is this critical flaw in the current methodology that this report seeks to address.

Drinking water regulations, federal and state, require that drinking water supplies meet compliance standards, MCLs, for a large number of analytes. The MCLs are generally based on available data regarding acceptable health risks for the wide range of contaminants regulated. While the uncertainties associated with these risk data are usually relatively large, MCLs are treated as “absolute” and “not to be exceeded” values. In addition, the analytical data developed in the analysis of drinking water samples collected for determining compliance with the regulations have varying degrees of analytical uncertainty. The relative degree of uncertainty for any analyte depends on a number of factors including: the method and instrumentation available for analysis; the laboratory performing the analysis; interferences from other contaminants, including natural and laboratory-generated contaminants; and factors associated with sample collection and shipping. Many of these factors are difficult to impossible to quantify, and much of the analytical data produced for compliance are not reported with even an estimate of the associated uncertainty, with radiological analyses being general exceptions. To further compound the problem, many MCLs are very near the minimum detection limit (MDL) or minimum reporting limits (MRL) for even the most sensitive instruments.

An earlier study<sup>2</sup> performed for the National Rural Water Association (NRWA) by SC&A, Inc. (SC&A) examined the analytical uncertainty associated with laboratory analyses of samples prepared with known analyte concentrations. These performance evaluation (PE) samples are prepared and analyzed for the purpose of evaluating laboratories certified by states and/or the EPA. The vast amount of data generated by the EPA WS study was used in the earlier study as the basis of evaluating analytical uncertainty for the analytes chosen for that study. This study introduced the concept of a Critical Value (CV) and calculated the CV for the ten analytes considered in that study. The CV is based on the confidence interval associated with analytical data at values near the MCL and was proposed as an alternative to the MCL for enforcement purposes. This is a more valid approach to the use of compliance data to eliminate unwarranted treatment options that, in fact, result in little or no health benefits.

The purpose of the present study was to obtain and evaluate drinking water compliance data from a representative number of states to assess the variability and reliability of these reported

---

<sup>2</sup> Reliability of Laboratory Data Generated for Compliance with Drinking Water Regulations, Submitted September 15, 2005.

data. Specifically the results of this study provide a representative determination of the statistical validity of using the mean of four-quarterly measurements for determining compliance with drinking water regulations.

## **2.0 Difficulty of Obtaining Compliance Data**

SC&A encountered unexpected difficulty in obtaining the raw data for the analyses performed for compliance with the Safe Drinking Water Act. We first attempted, on an informal basis, to get analytical data from a number of state programs with responsibility for administering the Act. After it became obvious that this would be unsuccessful, we made a concerted effort to obtain the data required for this project by formal means, i.e., written requests were submitted to those states that indicated usable data were obtainable. Additionally, we attempted to get assessable data from a website (<http://www.ewg.org/tapwater>) operated by the Environmental Working Group (EWG), a non-profit environmental and health related organization, as well as from National Contaminant Occurrence Database (NCOD)<sup>3</sup> maintained by EPA.

Drinking water data were finally obtained from the states of Alabama (2002-2004), California (Arsenic, 2002-2004), Michigan (2004-2005), Vermont (2005-2006), and Oregon (2002-2006). From these sources we extracted as much usable data as was reasonably possible for the purpose of our assessment. Information was also received from the states of Pennsylvania, Idaho, Utah, Rhode Island, and Missouri, but it was not usable because it only included incidences of MCL violations. While more limited in scope than we originally planned, we believe the data obtained and assessed are sufficient to raise concerns about the current procedures for enforcement of drinking water regulations.

The database maintained by EWG was not considered an acceptable source for data because of disparities in values obtained from this database as compared to those that should have been identical to ones obtained from the states. The NCOD, six-year review, database was in Access format and contained so many entries, most of which were designated as non-detects, that it was not efficient to extract the usable data from the vast amount of non-detects.

## **3.0 Study Approach**

Realistic assessments of water quality measurements can be made only if the statistical uncertainties inherent in monitoring measurements of the various analytes of interest are explicitly recognized. This study intended to employ statistically sound procedures for evaluating such measurement uncertainties for ten regulated analytes that account for most of the MCL violations, representing organic, inorganic, and radiological analyses. However, because of the limited amount of usable data that we were able to extract from the available sources, we abandoned our original plans to evaluate results for these ten analytes and chose data from any source and for any analyte for which we could positively identify four consecutive measurements from a given water system.

Due to the general practice of reporting the actual values obtained from radioactivity measurements, without regard to their relationship to reporting or detection limits, many of the

---

<sup>3</sup> Six Year Review Database

results of the statistical analysis in this study are for these types of measurements. This is contrasted against most of the data for other analytes where only “hits” or values above detection or reporting limits were supplied.

In the process of reviewing the data obtained, there were a significant number of instances where at least four results were reported for a given water system that were attributed to measurements made over a one- to two-day time period. The purpose for the repeated measurements is unknown, and it is also not possible to determine if the values represent multiple analysis of a single sample or the analyses of multiple samples. In either case, it is intuitively reasonable to assume that the variance for these analysis results, referred to herein as “short-term” or “repeated”, measurements, results would be less variant than the four-quarterly ones. Therefore, for purposes of comparison, we subjected these short-term (repeated) measurements to the same statistical analysis as the quarterly derived ones.

### 3.1 The mean is a variable

Consider the process currently used to determine compliance of a water system with respect to a given analyte. Four-quarterly compliance measurements, say  $x_1, x_2, x_3, x_4$ , are averaged and their mean  $\bar{x}$  is compared to the relevant MCL. If the mean does not exceed the MCL the water system is found to be in compliance, but, otherwise, the system is found to be noncompliant and remediation may be required, often at great cost to the water system and; therefore, to its customers.

The problem with this approach is that it completely ignores the highly statistical nature of the entire process. Consider the following issues:

- a) The quantities  $x_1, x_2, x_3, x_4$  are not (deterministic) numbers—they should be viewed as variables since, if they were redetermined (by repeating the sampling and laboratory processing procedures), it is almost certain that different values  $y_1, y_2, y_3, y_4$  would be obtained. Thus a different mean  $\bar{y}$  and, quite possibly, a different conclusion regarding compliance might well result.
- b) It follows then that the mean of quarterly compliance measurements should also be considered as a variable and that its variability should be explicitly considered when comparing the mean to the MCL.
- c) Of course the methods of probability and statistics offer tools for making such consideration of the uncertainty in the mean, but there is no definitive method for doing this. SC&A has developed two alternative procedures (“Method A” and “Method B”) and recommends both of them for consideration. Method B was detailed in the earlier aforementioned report prepared for NRW by SC&A, and this study represents Method A.

### 3.2 The confidence interval around the mean

One statistical parameter that has to figure prominently in any consideration of the

variability of the mean is the confidence interval for the mean. In constructing the standard confidence interval for the mean, one estimates the uncertainty,  $E$ , (as reflected by the data; typically  $E$  is determined at the 95% level of confidence) in the mean and then states the confidence interval as  $[\bar{x} - E, \bar{x} + E]$ , that is, the confidence interval consists of all values of concentration that are within  $E$  units of the mean. The specific value of  $E$  obtained for a given set of four-quarterly measurements depends on the confidence level chosen (we used 95% confidence throughout our work on this study) and the standard deviation. Ideally the standard deviation used would be determined from *all* possible measurements (the “*population* standard deviation”  $\sigma$ ) but this is almost never available for practical reasons. Therefore the predominant practice is to use the *sample* standard deviation,  $s$ , of the measurements  $x_1, x_2, x_3, x_4$ ; when this is done the uncertainty  $E$  takes the form:  $E = 3.182s / \sqrt{4} = 3.182s / 2$ .

#### 4.0 Results and Discussion

In this work, SC&A gathered a large amount of compliance data for many analytes and many water systems. Since the matter of interest here is the stability and reliability of the mean  $\bar{x}$  as a measure of the levels of various analytes in the water supplied by water systems, SC&A made comparisons, for each analyte and each water system, between the mean analyte level  $\bar{x}$  and its associated uncertainty  $E$  wherever there were sufficient data to allow this comparison to be made.

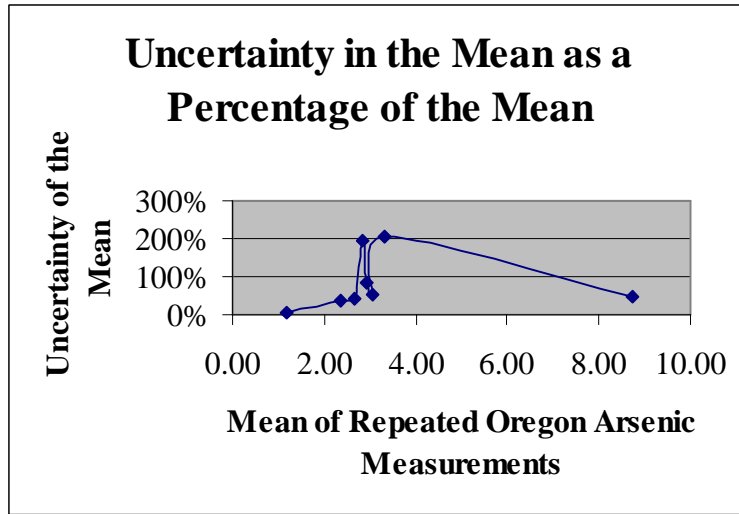
Graphs were constructed for those water supply systems and for those analytes where we were able to obtain at least two consecutive sets of four measurements. These graphs are contained in Appendix A. We believe that an elaborate quantitative analysis is not in order for this somewhat preliminary study and that a simple examination of the accompanying graphs will lead the reader to the appropriate conclusions as stated below.

The specific comparison mechanism used was to express  $E$  as a percentage of  $\bar{x}$ --this quantity is referred to as the “Uncertainty of the Mean” in the many graphs (of the uncertainty of the mean versus the mean, per se) and tables presented in the appendices, and a representative number below. As the reader examines these graphs and tables, several important points will become apparent:

- a) The uncertainty of the mean very frequently exceeds 100% of the mean, in other words  $E/\bar{x}$  is often greater than one—see the following graph of arsenic data from a water system in Oregon, for example. Notice that in this case, the uncertainty of the mean extends to over 200% in some instances.

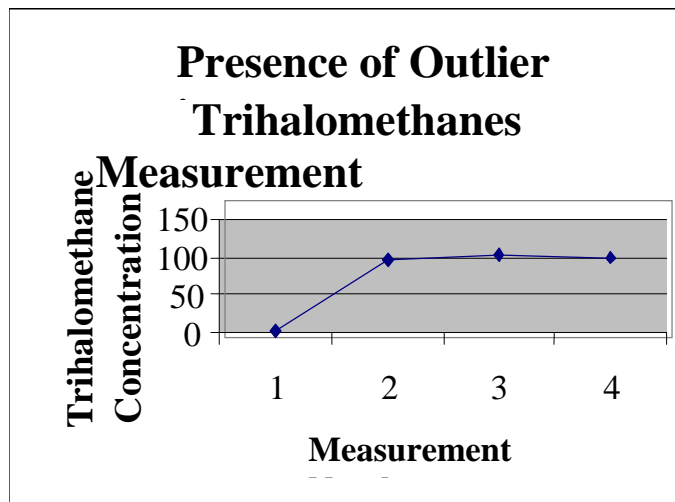
The importance of this is difficult to overstate. If one thinks of variables subject to statistical uncertainty as being “fuzzy,” then such an  $\bar{x}$  is so fuzzy that a distinction between the three concentrations (1) zero, (2)  $\bar{x}$ , and (3)  $2\bar{x}$  cannot be confidently drawn. Clearly, to make a decision to require remedial water treatment based on such

an  $\bar{x}$  value that has an attendant uncertainty level of 200% is highly questionable, if not illogical.



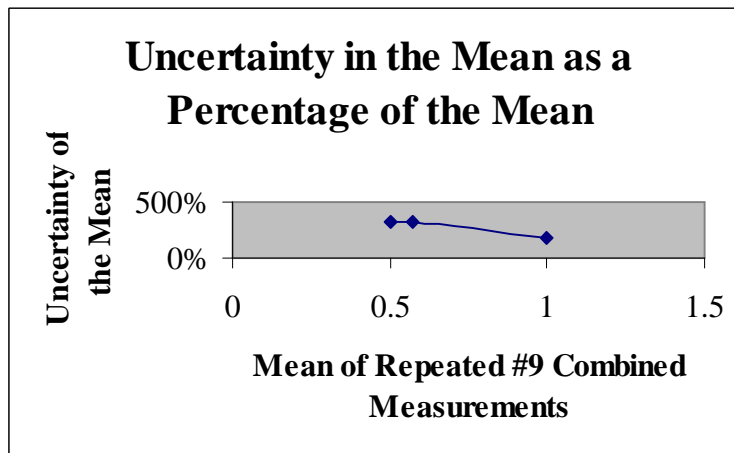
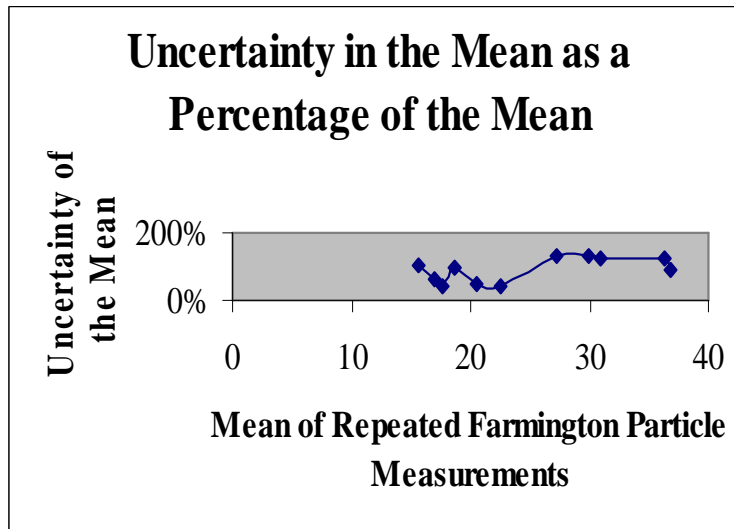
- b) Frequently, measurements that appear to be clearly outliers occur among the quarterly results. For example, consider the following consecutive values for trihalomethane concentration that were determined in one instance: (1) 2.89, (2) 95.3, (3) 102, (4) 99.6 Of course, it is possible that some environmental accident truly did cause trihalomethane concentrations to suddenly increase by two orders of magnitude, but, clearly, there is no sound rationale for averaging four numbers as disparate as these.

The presence of outliers greatly increases the standard deviation  $s$  and therefore also greatly increases the uncertainty  $E$  of the mean. SC&A believes that is very important that some type of outlier detection algorithm be devised and that resampling be conducted whenever a new quarterly measurement is suspected to be an outlier.





- c) Finally, it is interesting to compare the results for the quarterly data with those for the repeated measurements, i.e., those measurements that are made over a period of one to two days. One would certainly expect that the repeated measurements would exhibit generally lower uncertainty ratios than the quarterly data, but, while SC&A did not attempt to make any formal assessment of this issue, that does not seem to be the case at all: qualitatively, the repeated measurements appear to be about as highly scattered as the quarterly ones. See the following figures, for example.



This, of course, implies that the current sampling and sample processing procedures are often subject to very significant variances, a fact which strongly reinforces the contentions made in this report: at this time, there is far too much uncertainty inherent in the procedures used in making water quality assessments to ignore. SC&A urges that a statistical perspective of this entire enterprise be adopted.

## 5.0 Conclusions

In this work, SC&A gathered a large amount of compliance data for many analytes and many water systems. Since the matter of interest here is the stability and reliability of the mean of the

quarterly measurements as a measure of the levels of various analytes in the water supplied by water systems, SC&A made comparisons, for those analytes and each water system, between the mean analyte level and its associated uncertainty  $E$ , which SC&A determined using standard statistical methods. These comparisons were made, using graphical means, wherever there were sufficient data to allow such a comparison to be made.

The results of this comparative study revealed that the statistical uncertainty in the mean often exceeds the mean itself, i.e., the uncertainty of the mean is frequently more than 100% of the mean. The importance of this is difficult to overstate. If one thinks of quantities subject to statistical uncertainty as being “fuzzy,” then the mean of the quarterly measurements is frequently so fuzzy that a distinction between the three concentrations: (1) zero, (2) the mean, and (3) twice the mean cannot be confidently drawn. Clearly, to make a decision to require remedial water treatment action in the presence of such uncertainty is ill advised, if not illogical.

To assist the reader in visualizing the degree of uncertainty in four consecutive measurements, the following table was prepared, based on the data used to generate the graphs in Appendix A. The table presents the average uncertainty ratio ( $E/\bar{x}$ ), expressed as percentage, for each of the analytes represented in the graphs and the number of sets of four measurements (repeated or quarterly) whose uncertainty ratios are included in the averages shown here.

	<b>Arsenic</b>	<b>Copper</b>	<b>Fluoride</b>	<b>Gross Alpha</b>	<b>Ra-226</b>	<b>Ra-228</b>	<b>TOTAL HALOACETIC ACIDS (HAA5)</b>
Mean Uncertainty Ratio (%)	46.6	33.8	48.2	82.8	72.7	73.6	54.3
Number of Data Sets	30	5	4	80	23	19	20

Just as in the graphical representations, the high degree of uncertainty is obvious when presented in tabular form, as above.

There were a significant number of instances in the data we reviewed where multiple results were listed for a given water source and analyte with reporting dates separated by only one to two days. It is not known nor did we attempt to determine the reasons for these multiple short-term or repeated measurements. When we applied the same statistical analysis to these short-term (repeated) data as for the four-quarterly ones, there appeared to be no discernable difference in the variability (reliability) of the two. This was unexpected in that it is reasonable to believe that analyses for water samples collected and analyzed over a very short time frame would be less variant than those collected in four quarters over an annual time frame. This finding brings into question the usage of compliance data for enforcement purposes and suggests that the sampling and analytical processes may be incapable of supporting enforcement. In addition, it indicates that the practice of resampling a noncompliant system is no more likely to produce a reliable result than the original quarterly samples.

Surveying the data in the appendices reveals that, frequently, measurements that appear to be clearly “outliers” (outrageous values) occur among the quarterly results. Of course, the presence of outliers greatly increases the uncertainties determined from such data. SC&A believes that is very important to devise some type of outlier detection algorithm and that resampling be conducted whenever a new quarterly measurement is recognized as or suspected to be an outlier.

This study required the unnecessary expenditure of resources just to obtain drinking water compliance data in a format usable for even simple statistical manipulations. Given the importance of these data, the need for accessibility by reasonably skilled persons, and the mandates of the state and federal regulatory agencies, it seems almost negligent that they are not easily available from a regional or national database with easy access by interested parties.

The results of this study should be considered preliminary and the subject deserves additional study. Within the available resources, considering that a significant amount was spent to get the data, we were able to obtain sufficient data to examine the fundamental statistical foundation of compliance data. However, to assure that our preliminary findings are indeed representative, a much larger sample of data, extended to cover a greater geographical area and additional analytes, should be subjected to a more quantitative analysis.

## **APPENDIX A**

### **GRAPHS**

