A Study of the Precision of Lead Measurements at Concentrations near the Method Limit of Detection

Author(s): Paul M. Berthouex

Source: *Water Environment Research*, Vol. 65, No. 5 (Jul. - Aug., 1993), pp. 620-629
Published by: Wiley
Stable URL: https://www.jstor.org/stable/25044350
Accessed: 01-09-2020 22:12 UTC

A study of the precision of lead measurements at concentrations near the method limit of detection

Paul M. Berthouex

ABSTRACT: Many problems in environmental quality require measurements on chemicals that are expected to be absent or to exist at very low concentrations. Test specimens delivered to the laboratory for analysis will contain literally "next to nothing" which may lead to some observations being reported as "below the limit of detection." The limit of detection is a statistical concept that is intended to reflect the magnitude of unavoidable random fluctuations in measurements at these low concentrations.

The purpose of this study was to generate a collection of measurements on samples that have low concentrations, at or near the limit of detection. The results show the magnitude of measurement variability at low concentrations and have some important implications regarding the utility of the limit of detection concept in compliance monitoring.

An interlaboratory study involving eight laboratories, each of which measured lead on 50 test specimens, provides a large collection of measurements on lead at low concentrations, including concentrations below the laboratories' stated limits of detection. The data show that the variability of measurements at low concentrations was not greater than at the higher concentrations. Lead was not only detected in low concentration specimens, it was also quantified, even at concentrations near and below the limit of detection as it is usually defined, especially when data are interpreted collectively instead of as isolated individual values. The accuracy of measurements may be affected more by bias than by poor precision.

The method detection limit (MDL) is a misunderstood and often misused concept. The concept of an MDL does not deal with bias; it is only concerned with precision. It is a quantity that cannot be estimated with great precision. Throwing away measured values because they are "below the limit of detection" discards much useful information and it is recommended that this practice be discontinued. *Water Environ. Res.*, **65,** 620 (1993).

KEYWORDS: bias, lead, limit of detection, monitoring, precision.

Many important scientific problems in environmental quality focus on chemicals that are expected either to be absent or to exist at very low concentrations. The test specimens delivered to the chemist for analysis will contain literally "next to nothing." Under these conditions it can be expected that many data sets will include some observations that are reported only as "below the limit of detection." The limit of detection is a statistical concept that is intended to reflect the magnitude of unavoidable random fluctuations in measurements at these low concentrations. We assume that the detection limit of interest is the *method detection limit* (MDL).

The purpose of this study was not to compare the performance of different laboratories, although to some extent the data do lend themselves to doing this. The objective was to generate a collection of measurements on samples that have concentrations near or below the limit of detection. The results will be used to show the magnitude of measurement variability at low concentrations and to evaluate the utility of the limit of detection concept in monitoring.

Comments About the MDL

The concept of using an MDL has become so widely accepted that some important misconceptions surround its use. Three are briefly discussed.

The MDL is a statistical concept. The MDL is often considered to be a chemical concept. It is not. It is a statistical concept. The limit of detection is generally defined as "the lowest concentration that can be determined to be statistically different from a blank specimen." In order to apply this definition it is necessary to know, or estimate, the average measurement error of blank specimens. How the chemist operationally determines this estimate is extremely important (see, for example, Am. Chem. Soc., 1983; Glaser *et al.*, 1986; Holland and McElroy, 1986; Hunt and Wilson, 1986; and Kaiser and Menzes, 1968), but this will not be discussed here.

The MDL is an imprecise quantity. Many people have come to believe that the MDL is a fixed and known quantity. It is not. The MDL is a statistic that is estimated from data, just as an average is a statistic that is estimated from data. I have never actually heard anyone say "the atomic weight of lead is 207.19 and its MDL is 2.1," but it is common for chemists, engineers, and managers to discuss the MDL as though it were a known reference value like atomic weight. Unfortunately, it is a fuzzy value that differs from laboratory to laboratory, effluent to effluent, analyst to analyst, and so on. As an example, suppose that the true (but unknown) standard deviation of blanks is 1.0 and, furthermore, that we have made measurements on many sets of replicate blanks, and from each set estimated a standard deviation. A wide range of values would be found in this case. For example, if each set of replicates consisted of eleven blanks, 95% of the estimated standard deviations would be expected to fall within the range of 0.52–1.93. In terms of estimating the MDL, this means that the estimated MDL = 2.76 on average, and that 95% of the estimates would fall within the range of MDL = 1.5 to MDL = 5.0. That is, the 95% confidence interval on the MDL would be 1.5–5.0. It would be expected that 5% of the time, in many repeated estimates, that an estimate outside this range would be obtained for the value that we call the MDL. We see that the MDL is not a value that can be known with a great deal of confidence.

The MDL says nothing about bias in measurements. The MDL is often treated as a measure of the accuracy of low-concentration test results. Accuracy is a combination of fidelity to a true value

and the variability of repeated measurements about this true value. *Accuracy* is a function of both *bias* and *precision.* Bias measures systematic errors and precision measures the degree of scatter in the data. Methods that give accurate measurements have good precision and near zero bias. Inaccurate methods can have poor precision, or unacceptable bias, or both.

The MDL reflects only variability. It says nothing about bias, or the lack of it. We will see later that measurements can have low variability but still show considerable bias with respect to the true concentration.

Random errors affect the precision of a measurement. Precise results have small random errors and the scatter between repeated measurements is small. Random errors can never be eliminated, though by careful technique they can be minimized. Their effect can be reduced by making repeated measurements and averaging them. Making repeated (replicate) measures also provides the means to quantify the measurement errors and evaluate their importance. Bias cannot be averaged out by making more measurements.

Reproducibility and *repeatability* are often used as synonyms for precision, but a distinction between them should be made. Suppose an analyst made the five replicate measurements in rapid succession, say within an hour or so, using the same set of reagent solutions and glassware throughout. Temperature, humidity, and other laboratory conditions would be nearly constant. Such measurements would estimate repeatability, which might also be called within-run precision. If the same analyst did the five measurements on five different days, the results would reflect additional variation due to differences in laboratory conditions, reagents, and other factors. These data would indicate the reproducibility of the measurement process. It would be a misrepresentation to report repeatability as reproducibility.

Problems with bias and poor precision are not hypothetical problems, as shown by the actual measurements on lead in a wastewater effluent matrix plotted in Figure 1. Eight laboratories each made six measurements on lead in filtered unspiked wastewater effluent. Three samples were done in one batch and another three samples, identical to the first three, were analyzed later. Laboratory 3 has good precision, but seems to be biased low. Laboratory 1 has poor precision and is biased high. Laboratory 7 has a lot of variability but its average compares well with most of the other laboratories. Laboratory 2 has good precision within each batch of three specimens, but one batch is biased low.

Reporting Low Concentrations

There seems to be different practices among chemists in the way observations near and below the limit of detection are reported. They may report the datum to the data analyst as (1) trace, (2) the letters ND (not detected), (3) the numerical value of MDL itself, (4) a "less than" value, that is, the numerical value of the MDL preceded by a "<" sign, (5) zero, (6) some value between zero and the MDL, for example, one-half the MDL, (7) the actual measured concentration even if it is below the MDL (that is, whether the value is positive or negative), (8) the actual measured value followed by the MDL in parenthesis, or (9) the actual measured value with a statement of its precision (for example, 2 ± 4 μg/L, where the ± value indicates the precision of the estimate). The last three methods are the best (Gilbert, 1987; Hunt and Wilson, 1986; and Rhodes, 1981).

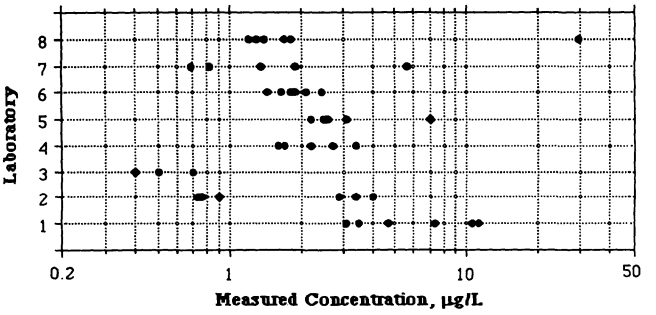

Figure 1—Replicate measurements on filtered unspiked effluent made by eight laboratories.

The way data are reported may determine whether evidence regarding the accuracy of the measurement process can be presented in court proceedings. Koorse (1991) stated that in some 50 cases to date, the courts have not allowed argument that the deviation from a standard was within the error band of the measurements on which the decision was based. The reason was not the court's failure to understand the concept of measurement variation, but rather that the defendant's monthly reports were signed as being an accurate reflection of the quality of the effluent. Signing such a statement negated the possibility of later claiming that the measurements were affected by measurement error. Koorse (1991) suggested that wastewater utilities might attach a general statement to each report stating that the values reported are estimates of the true concentrations and as such they reflect variations in the sampling and measuring process. An alternative to a general exculpatory statement would be to report each value, whatever it might be, along with a statement of its precision. This would clearly indicate that the reported measurements are estimates, and can reflect changes in precision of the measurement process that are bound to occur over time.

Study Guidelines

Five treatment plant laboratories, two commercial laboratories, and the Wisconsin State Laboratory of Hygiene participated in the study. All laboratories used the approved EPA atomic absorption graphite furnace method for lead in an aqueous solution.

The test specimens were prepared from an activated sludge effluent that was filtered to remove particles. A sufficient quantity of this filtered effluent was prepared to make the test specimens that were needed, plus enough to provide each laboratory with a 1 L portion of "blank matrix" with the first batch of samples.

All test specimens were prepared at the same time. They were randomly assigned to batch and laboratory. Each laboratory received two batches of 25 test specimens, for a total of 50 that were distributed as follows: 6 blank (unspiked) specimens, 20 specimens with a 1.25 μg/L spike, 14 with a 2.5 μg/L spike, 5 with a 5 μg/L spike, and 5 with a 10 μg/L spike. These were apportioned almost equally between the two batches.

There was no way to disguise the fact that the test specimens were part of a special study, but the laboratories were kept blind to certain key information about the specimens. They were told that the study was about measurements on "low" concentrations, but they were not told what range of lead concentrations to

expect, how many different concentrations were represented, that some specimens were unspiked, or any other information that could influence the values they reported.

The key instructions that were given to each participating laboratory were:

- Do not try to make especially careful measurements for this study.
- The test specimens were all prepared from the same background matrix, which was filtered final effluent from an activated sludge treatment plant that produces high quality effluent with a naturally low concentration of lead. A 1-L container of blank background matrix has been provided in case your laboratory wishes to establish the MDL using the actual background matrix.
- The test specimens have been spiked with known amounts of lead, added in a soluble form, and preserved with nitric acid.
- Do not report any results as "less than MDL." Report a numerical value for each sample, even if the value is zero, negative, or below your MDL.

Results of the Interlaboratory Study

Table 1 summarizes the average and standard deviation at each concentration level for each laboratory. Laboratory 1 is biased high, and also has poor precision. Laboratory 2 measures at approximately the same level regardless of the lead concentration of the test specimen. Because of these obvious difficulties, general statements about the lead measurement process will be based on the performance of laboratories 3 through 8. One observation is that the variability of measurements is not larger at the lowest concentrations.

Figure 2 shows the measured values for each laboratory plotted against the amount added to the specimens. The solid line shows the result that would be obtained if the measured concentration equaled the added (spiked) concentration. Since the background concentration was not zero, the results should not fall on this line, but it nevertheless serves as a useful reference. We expected the measured concentrations to be consistently offset above this line, such as the results shown for laboratories 4, 5, and 6. Laboratories 3, 7, and 8 are above the line at low concentrations, but on it or below at high concentrations, which indicates underestimation at the higher levels.

Figure 3 summarizes the results of all measurements in a form that shows the spread of the values obtained by each laboratory at each lead concentration. Figure 4 is a comparison of the measurements on the unspiked, 1.25 μg/L, and 2.5 μg/L specimens for all laboratories. This plot is scaled to show only those values below 15 μg/L (approximately 10 high values were omitted). A log scale was used to facilitate showing the extremely high values. Unspiked results are indicated by a cross (+) on the left, the 1.25 μg/L spike results by a diamond in the center, and the 2.5 μg/L spike results by open circles on the right. The 48 measurements on unspiked matrix yielded values ranging from 0.3 μg/L to 30 μg/L, with more than half the values falling into the range of 1–3 μg/L.

The between and within laboratory variability makes it difficult to state the background concentration with a high degree of confidence, but a useful estimate is 1.6 μg/L. It appears that the 1.25 μg/L spike specimens, on average, are approximately 1.25 μg/L above this level, and that the other spikes have shifted the measured results by about the right amount. This shows a rather useful degree of consistency. Even at these low concentrations, the analyte was detected and repeatable numerical values were measured. The three levels of lead not only were detected, they were well quantified by six of the eight laboratories. The spread of values at the three concentration levels is about the same for all the laboratories. The 1.25 μg/L difference between concentrations is evident. Considering each laboratory individually, it would appear that their measurement process is under control, in the sense that the 2.5 spikes are higher than the 1.25 spikes which are higher than the unspiked results. Also, the variation on the log scale is approximately the same at the three levels. This consistency shows that laboratories are capable of producing useful information about low concentrations.

Another significant impression is that the variability within a particular laboratory is random. This means that the variation

Table 1—Summary statistics for the eight laboratories measuring lead at five different concentration levels.

	Laboratory							
Specimen and statistic	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Unspiked (n = 6)								
Average, μg/L	6.7	2.1	0.6	2.3	3.3	1.9	1.8	1.5
Standard deviation, μg/L	3.5	1.5	0.1	0.7	1.9	0.4	1.9	0.3
1.25 μg/L spike (n = 20)								
Average, μg/L	6.7	1.9	1.2	2.7	3.3	2.8	2.1	2.2
Standard deviation, μg/L	2.8	0.9	0.3	1.1	1.1	0.5	1.2	2.0
2.50 μg/L spike (n = 14)								
Average, μg/L	8.9	1.7	2.7	4.6	4.7	4.5	3.1	3.2
Standard deviation, μg/L	4.0	0.9	0.3	1.1	1.8	0.6	0.6	0.8
5.00 μg/L spike (n = 5)								
Average, μg/L	28.2	2.0	5.1	8.1	6.3	6.9	3.4	5.7
Standard deviation, μg/L	27.1	1.1	0.1	1.5	3.3	0.7	2.1	0.3
10.0 μg/L spike (n = 5)								
Average, μg/L	19.2	3.1	9.4	12.3	21.8	11.2	9.1	10.7
Standard deviation, μg/L	11.0	0.2	0.2	2.1	16.9	1.1	1.1	1.2

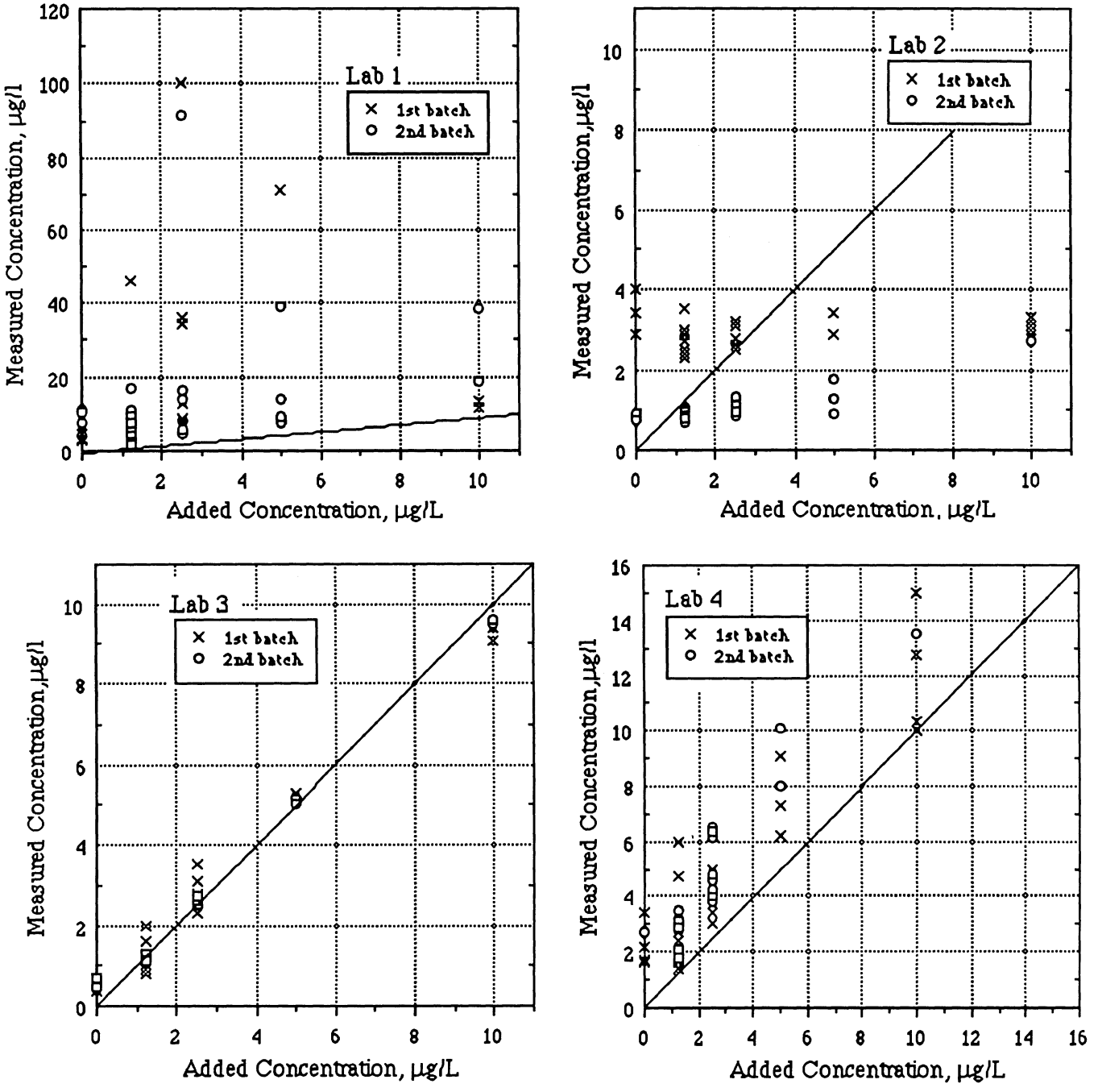

Figure 2—Results measured on the two batches of test specimens.

can be averaged out by making repeated measurements. That is, while individual values are only semiquantitative, a collection of data can have a clear meaning. This would mean that variability is not an insurmountable difficulty in low-concentration specimens.

Bias, rather than precision, may be the problem of greatest importance. There are differences between laboratories that are not only statistically significant—they are large.

Estimating the Limit of Detection

It is interesting to see what MDLs are calculated using the EPA definition (MDL = st where t is the Student's t value for a one-sided 99% confidence level and s is the standard deviation estimated with $n - 1$ degrees of freedom, where $n \geq 7$).

Tables 2 and 3 show that widely different values can be estimated in different laboratories. Even within a single lab, the estimated value can vary considerably depending on which test results are used. From Figure 5 it would appear that the MDLs for the most consistent laboratories are in the range of 1.5–3.0 µg/L (note that the largest values were omitted from this plot).

Agreement between measurements on replicate specimens of unspiked matrix that were prepared in the participant laboratory was much better than agreement between measurements on blind specimens of the same material. This does not suggest intentional

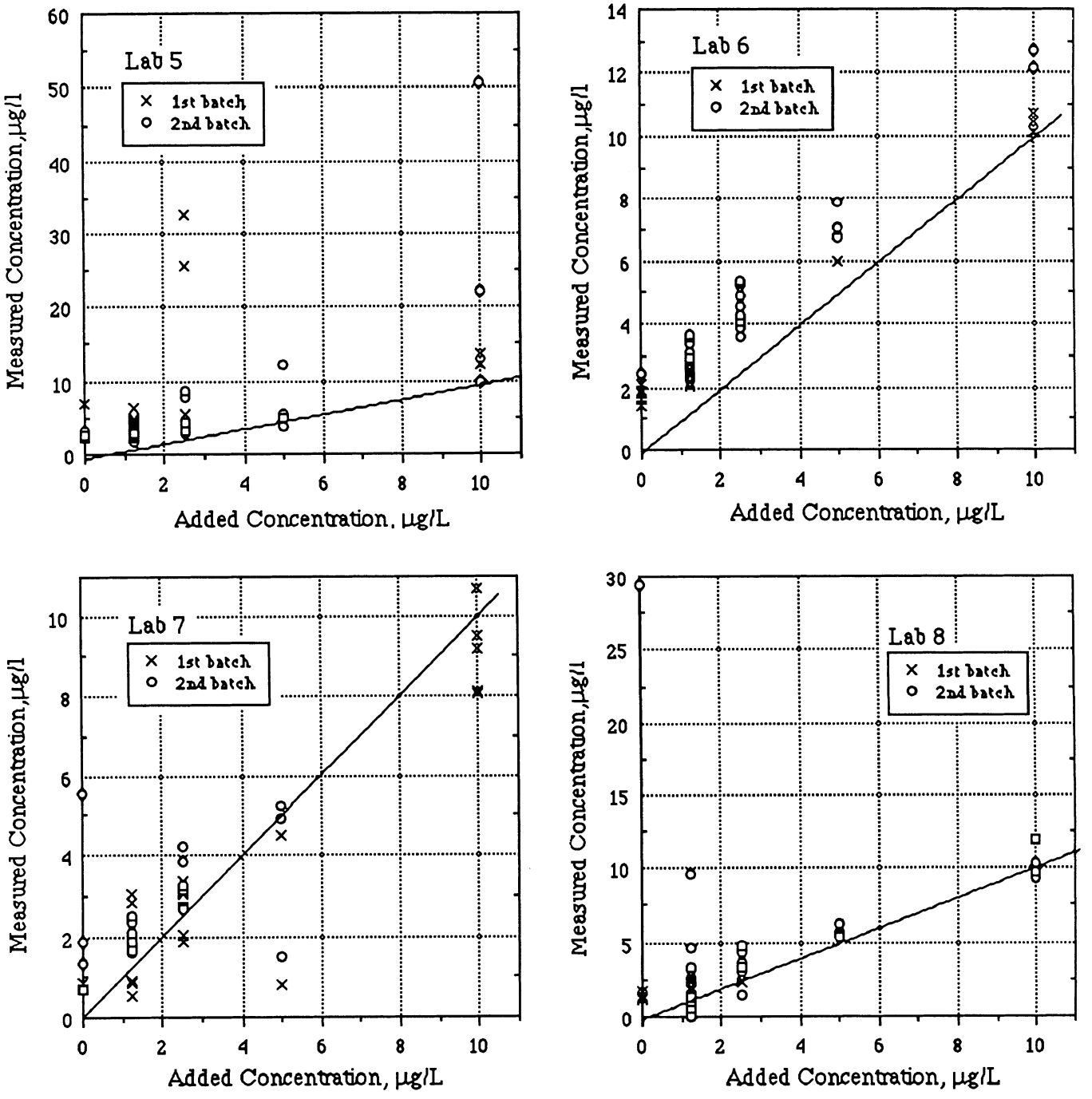

Figure 2—(*continued*) Results measured on the two batches of test specimens.

bias on the part of the participating laboratories, but it may suggest an inherent difficulty when samples are known to be alike.

Implications for Monitoring

Data produced by a real system will be variable because of random measurement error, like that which is so apparent in the lead measurements shown in this paper, and also real variation in the underlying true state of the system. *Random error* refers to unavoidable variation due entirely to the sampling and measurement process that cannot be eliminated or reduced by changing the true quality of the effluent. Bias is another possible source of variation in the measured values (for example, laboratory 1 was consistently too high), which labs expect to eliminate by careful technique, but it is always difficult to guarantee that this has been accomplished unless external quality control checks are used.

In a monitoring situation we must use the available data to make decisions about the true state of the system, such as whether it is in compliance with a legal standard. The primary difficulty in doing this lies in knowing how to characterize and recognize random variability so it is not interpreted as real change. This

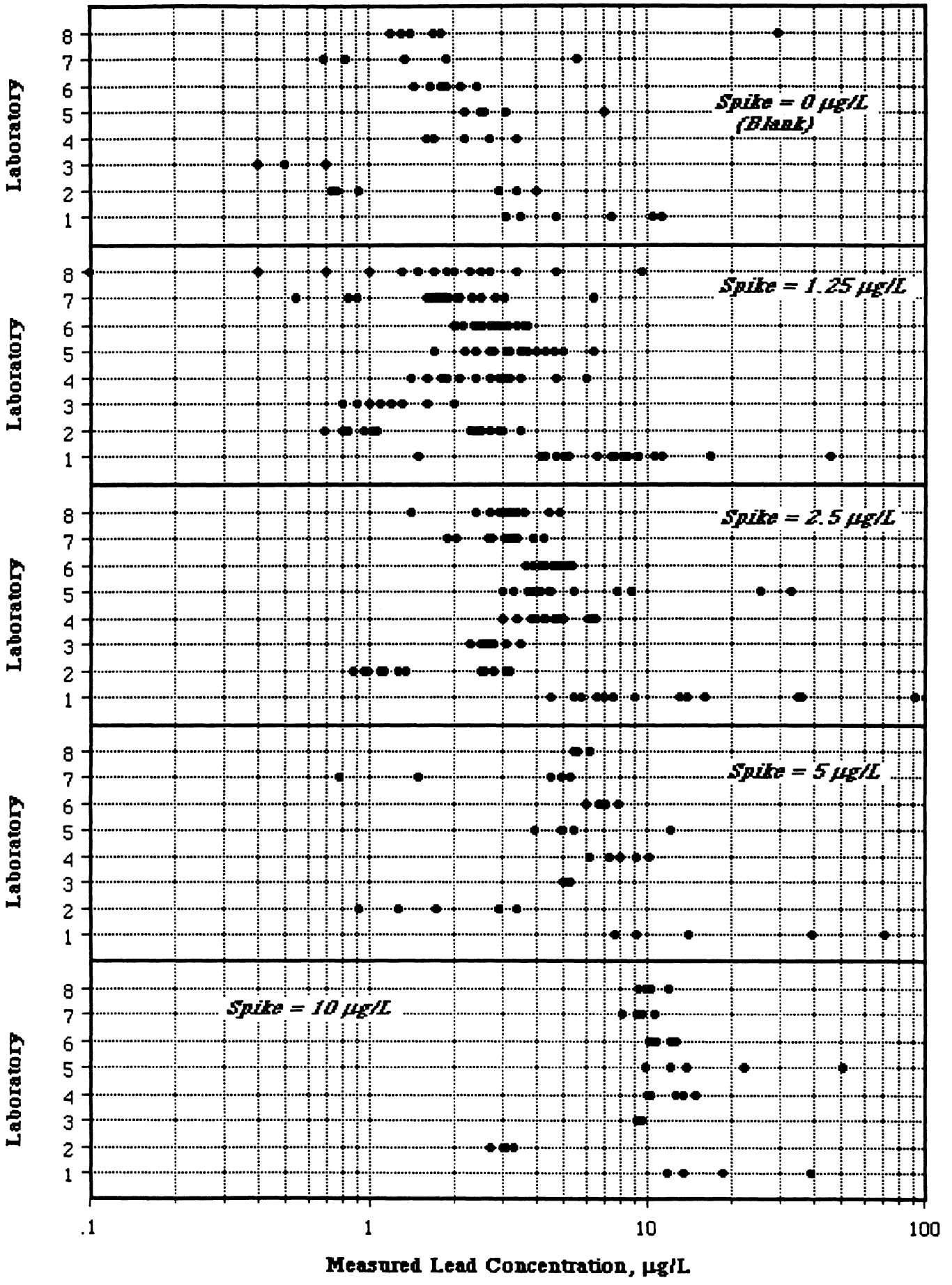

Figure 3—Comparison of all measurements by the eight laboratories.

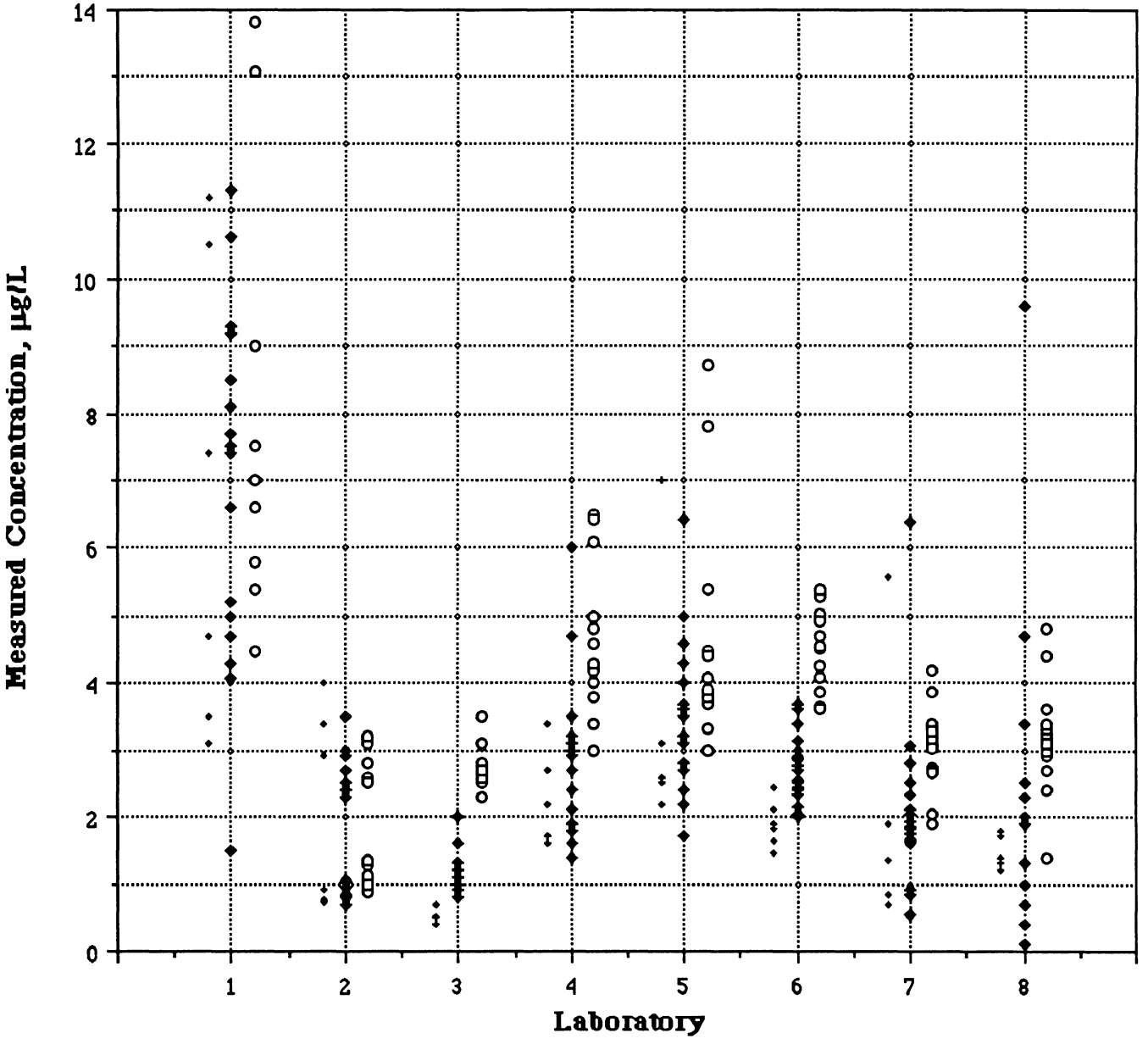

Figure 4—Comparison of the unspiked, 1.25 µg/L, and 2.5 µg/L spiked results that are less than 15 µg/L (nine higher values are not shown). Unspiked results are indicated by a cross (+), the 1.25 µg/L spike results by a diamond, and the 2.5 µg/L spike results by open circles.

difficulty is illustrated by constructing a time series of effluent concentrations from values reported on unspiked and 1.25 µg/L spiked matrix. The true concentration of all the tested specimens is certainly always less than 3 µg/L (the unspiked samples have a true concentration less than 2 µg/L). Values greater than 5 µg/L were not plotted. This "time series" is shown in Figure 6. The bold horizontal line is drawn to represent an MDL of 2.5 µg/L, which would be a reasonable value for a reliable laboratory.

The graph illustrates two points. First, notice how many measurements would have been lost if an MDL had been applied at the point of data generation. Most of the record would disappear; only the part above the shaded area would be available for interpretation. This loss of useful information would strongly affect our impression about the quality of the effluent.

Second, we almost always underestimate the true magnitude of random measurement error. The great investment of effort, talent, and money in making measurements creates a natural reluctance to admit imperfections in methods. Unfortunately, this underestimation has serious consequences. It will result in wasting money to investigate and "cure" imaginary problems. We illustrate this with the constructed time series shown in Figure 6.

Suppose that the effluent must satisfy a standard of 3 µg/L. We know from the origin of the samples that every specimen tested was in compliance with this standard. That is, the effluent

Table 2—Estimate of the method limit of detection (referenced to zero) using unspiked specimens.

	Laboratory							
Unspiked specimen	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Blind replicate	4.70	4.00	0.7	2.2	2.5	1.45	0.83	1.8
Blind replicate	3.50	2.90	0.5	1.7	7.0	2.12	0.69	1.2
Blind replicate	3.10	3.40	0.4	3.4	2.2	1.90	1.89	1.3
Blind replicate	11.20	0.73	0.5	1.6	2.2	1.81	1.35	1.4
Blind replicate	7.40	0.91	0.7	2.2	3.1	1.65	0.69	1.7
Blind replicate	10.50	0.77	0.7	2.7	2.6	2.45	5.57	29.4[a]
Laboratory replicate	23.60[a]	0.80	−0.4	4.5	1.3		1.33	2.5
Laboratory replicate	4.40	0.80	0.1	3.3	0.5		2.26	3.2
Laboratory replicate	4.00	0.69	−0.3	2.9	0.5		2.89	3.5
Laboratory replicate	3.60	0.91	−0.4	4.3	1.2		1.47	3.3
Laboratory replicate	4.10	0.98	−0.5	2.5	1.6		1.35	3.1
Laboratory replicate	0.98		−0.1	3.3	1.1		0.98	2.8
Laboratory replicate			−0.4	2.5	1.0		1.10	4.0
Laboratory replicate			−0.1		2.9		0.79	
Laboratory replicate			−0.5					
Laboratory replicate			−0.7					
Blind unspiked specimens								
n =	6	6	6	6	6	6	6	6
Average, μg/L	6.73	2.12	0.58	2.30	3.27	1.90	1.84	1.48
Standard dev., μg/L	3.52	1.48	0.13	0.67	1.86	0.35	1.89	0.26
MDL, μg/L	11.8	4.98	0.44	2.25	6.26	1.12	6.36	0.87
Laboratories' replicates								
n =	5	6	10	7	8	0	8	7
Average, μg/L	4.02	0.86	−0.33	3.33	1.26	—	1.52	3.20
Standard dev., μg/L	0.33	0.17	0.24	0.80	0.76	—	0.71	0.48
MDL, μg/L	1.49	0.23	0.67	2.51	2.28	—	2.13	1.51
All values								
n =	12	11	16	13	14	6	14	13
Average, μg/L	5.17	1.49	0.12	2.85	2.112	1.89	1.66	2.48
Standard dev., μg/L	2.99	1.20	0.50	0.89	1.64	0.35	1.29	0.38
MDL, μg/L	9.12	3.33	1.30	2.38	4.35	1.12	3.42	1.02

[a] Treated as an outlier.

is truly in compliance 100% of the time. Nevertheless, because of random measurement errors, values greater than 3 μg/L were measured 22 times out of 120 observations. It would be a waste of resources if these 22 occasions were declared violations. Our approach to monitoring, including the writing of performance standards, must be altered if such mistaken judgments are to be avoided.

Implications Regarding How Data are Reported

The instruction to all participating laboratories to "Report a numerical value for each sample, even if the value is . . . below your MDL" produced an interesting example of data censoring. All laboratories but one complied with this instruction. The exception (laboratory 6) reported numerical values for less than ten specimens. Fortunately, the raw output of their instrument was still available so measurements for all 50 specimens could be recovered; the "lost data" were the most consistent of any laboratory. The widespread practice of censoring data (that is, reporting only those values above the detection limit) almost certainly wastes a tremendous amount of useful information. Perhaps this anecdote can help persuade laboratories to start

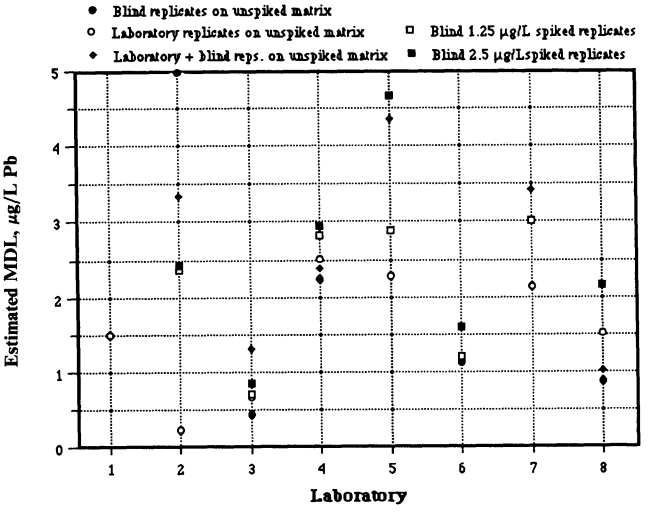

Figure 5—Various estimates of the MDL computed using data from the eight participating laboratories. Several estimates larger than MDL = 5 were omitted in making this plot.

Table 3—Estimates of MDL (μg/L) using 1.25 μg/L and 2.5 μg/L spiked specimens.

	Laboratory							
Specimen	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1.25 μg/L spike	9.2	3.00	1.1	6.0	2.8	2.35	1.93	1.7
(n = 20)	8.5	2.70	2.0	1.9	2.7	2.86	0.54	1.9
	10.6	2.30	1.3	2.9	6.4	2.70	0.84	1.7
	8.1	2.70	1.0	1.8	2.4	2.56	0.90	2.7
	4.3	2.40	1.1	1.4	5.0	2.88	3.06	2.0
	45.8[a]	2.30	0.8	1.6	3.7	2.04	2.82	1.5
	5.0	2.50	0.8	2.7	4.6	2.78	1.93	1.9
	7.4	2.30	0.9	2.4	3.1	2.16	2.03	0.1
	7.5	2.90	0.8	4.7	3.6	2.43	1.61	2.3
	11.3	3.50	1.6	1.8	3.1	2.42	1.67	2.5
	9.3	2.50	1.1	1.9	4.3	2.01	1.69	1.9
	4.1	1.02	1.2	3.2	4.0	2.50	2.10	4.7
	7.4	1.06	1.3	3.1	1.7	2.98	1.87	0.7
	5.2	0.80	1.2	3.2	2.2	2.87	6.38	0.4
	16.7[a]	1.02	1.3	3.0	2.4	3.39	2.34	1.0
	4.7	1.02	1.1	3.5	3.5	3.60	1.82	3.4
	6.6	0.95	1.2	3.1	2.2	2.99	1.75	0.4
	1.5	0.69	1.2	1.8	2.7	3.68	1.85	1.3
	7.7	0.84	1.3	2.1	3.2	3.13	2.53	2.0
	1.5	0.80	1.1	2.9	2.8	2.90	1.63	9.6
Average, μg/L	6.66	1.86	1.17	2.75	3.32	2.76	2.06	2.18
Standard dev., μg/L	2.76	0.93	0.28	1.11	1.12	0.47	1.18	2.05
MDL, μg/L	7.12	2.37	0.71	2.83	2.88	1.20	3.01	5.23
2.5 μg/L spike	100.0[a]	2.80	2.8	3.4	4.5	5.30	3.15	3.3
(n = 14)	13.1	2.60	3.5	3.0	3.7	4.72	3.39	2.4
	35.9[a]	3.10	2.3	3.4	3.8	3.64	3.08	2.7
	34.6[a]	2.50	2.7	5.0	32.6[a]	5.04	3.04	3.2
	6.6	3.20	2.3	6.1	5.4	3.62	3.88	3.3
	5.4	1.09	3.1	6.5	3.9	4.53	2.04	2.9
	7.0	1.27	2.5	6.4	4.1	4.57	1.89	4.4
	9.0	1.35	2.5	4.2	25.6[a]	5.27	2.74	3.4
	7.5	1.13	2.5	3.8	3.0	3.88	3.85	4.8
	16.2	0.95	2.7	4.2	4.5	4.10	4.20	3.0
	13.8	0.98	2.5	4.6	7.8	4.25	3.16	3.6
	4.5	0.87	2.5	4.8	3.3	4.26	2.71	3.1
	5.8	0.87	2.6	4.0	8.7	5.39	2.67	3.4
	91.3[a]	0.98	2.7	4.3	4.4	4.91	3.27	1.4
Average, μg/L	8.9	1.69	2.66	4.55	4.71	4.53	3.08	3.21
Standard dev., μg/L	4.04	0.91	0.32	1.11	1.76	0.60	0.65	0.81
MDL, μg/L	10.76	2.41	0.85	2.94	4.66	1.59	1.73	2.16

[a] Treated as outliers.

reporting the values as measured (which can be accompanied by an appropriate statement about their precision).

Summary

Measurements on test specimens containing lead at low concentrations were provided by eight laboratories. These data show that the measurement process can produce useful numerical data even when concentrations are at and below the limit of detection as it is usually defined. Lead was not only detected in low concentration specimens, it was also quantified.

A weakness in monitoring is an insistence on looking at single values in isolation from other available data. This practice defies common sense and is also weak on scientific grounds. Standards and decisions that rest on judging a single value, no matter what measurement system produced the value, are inherently weak. Perhaps it is recognition of this weakness that has strengthened the tendency toward reporting "<MDL" as a means of discounting single values. This is the wrong solution. A better approach is to look collectively at all available information so that small random errors are averaged out and the bigger picture can be seen.

The MDL is often misunderstood and misused. Its use may cause more problems than it solves. It is not a quantity that can

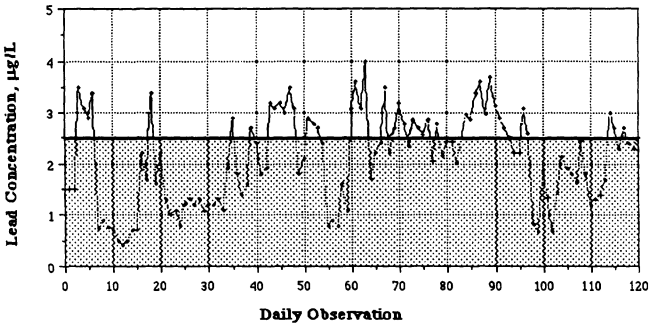

Figure 6—Data series constructed from the measurements reported on unspiked matrix and 1.25 µg/L spiked matrix. Values above 5 µg/L were omitted. The bold horizontal line represents a "reasonable" MDL of 2.5 µg/L.

be estimated with great precision, perhaps not even with good precision.

Not reporting numerical values discards much useful information and it is recommended that the practice of censoring the numerical values at the point of data generation be ceased.

The accuracy of measurements may be affected more by bias than by poor precision. There were significant differences in the levels measured by different laboratories. The concept of an MDL does not deal with bias; it is only concerned with precision.

Acknowledgments

Credits. Paul Fritschel, manager of the Environmental Engineering Laboratory of the Department of Civil and Environmental Engineering, The University of Wisconsin-Madison, provided valuable advice and supervised the preparation of the test specimens. Diane Fiegel prepared the test specimens, coordinated preparation of the database, and did some preliminary analysis of the data.

We thank the sponsors of the Wisconsin Consortium for Applied Water Pollution Research for their financial support and for their laboratories' cooperation in making the lead measurements reported in this paper. We also thank Green Bay, Janesville, and the State of Wisconsin Laboratory of Hygiene for their participation in the study.

Authors. Paul M. Berthouex is a professor in the Department of Environmental Engineering, The University of Wisconsin-Madison. Correspondence should be addressed to Professor Paul M. Berthouex, Dept. of Civil and Environmental Engineering, University of Wisconsin, 1415 Johnson Dr., 3204 Engineering Building, Madison, WI 53706.

Submitted for publication March 4, 1992; revised manuscript submitted November 9, 1992; accepted for publication November 10, 1992. The deadline for discussions of this paper is November 15, 1993. Discussions should be submitted to the Executive Editor. The authors will be invited to prepare a single Closure for all discussions received before that date.

References

American Chemical Society (1983) Principles of Environmental Analysis. *Anal. Chem.*, **55,** 2210.

Gilbert, R. O. (1987) Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold Co., New York.

Glaser, J. A., Foerst, D. L., McKee, G. D., Quave, S. A., Budde, W. L. (1981) Trace Analyses for Wastewaters. *Environ. Sci. Technol.*, **15,** 1426.

Holland, D. M., and McElroy, F. F. (1986) Analytical Method Comparison by Estimates of Precision and Lower Detection Limit. *Environ. Sci. Technol.*, **20,** 1157.

Hunt, D. T. E., and Wilson, A. L. (1986) The Chemical Analysis of Water. 2nd ed., The Royal Society of Chemistry, London, G.B.

Kaiser, H. H., and Menzes, A. C. (1968) Two Papers on the Limit of Detection of a Complete Analytical Procedure. Adam Hilger Ltd, London, G.B.

Koorse, S. J. (1991) Clean Water Regulatory Issues: A Legal Perspective. Keynote Address, 64th Annual Meeting of the Central States Water Pollut. Control Assoc. St. Charles, Ill., May 14–17.

Rhodes, R. C. (1981) Much Ado About Next to Nothing, Or What to Do with Measurements Below the Detection Limit. Environmetrics 81: Selected Papers, SIAM-SIMS Conference Series No. 8, Philadelphia, Pa., 157.