An Informal Assessment of Interlaboratory Variability of Mining Effluents

Derek Riehm, P.Eng.
Barrick Gold Corporation, Lima, Perú
(formerly with Teck Cominco Limited)

Guy Gilron, M.Sc., R.P.Bio.
Cantox Environmental Inc., Mississauga, Ontario

Mark Thompson
Teck Cominco Alaska Incorporated, Anchorage, Alaska

Abstract

How variable are effluent toxicity testing data? One common response to this question is that the variability of toxicity tests is comparable to the variability associated with analytical chemistry data. For toxicity testing, this is based in large part on data derived from reference toxicant tests employing analytical-grade solutions. What about the variability associated with toxicity of real-world mining effluents, which are to be tested in the upcoming metal mining environmental effects monitoring (EEM) program in Canada? Few interlaboratory data sets using real mine effluents currently exist. In this presentation, we review available data comparing analytical and toxicological data variability among laboratories, using the relative percent difference (RPD) as a yardstick. Recognizing that the method detection limit (MDL) used in analytical variability lacks an obvious counterpart in toxicity testing, we found that toxicity test interlaboratory variability is considerably greater than analytical variability, in particular for analyses conducted at concentrations well above the MDL. Interlaboratory variability is important where toxicity test results are used to demonstrate compliance or where they may be used, for example, in an EEM program to trigger follow-up action. Our assessment demonstrates the importance of studying this issue further. An understanding of the issue of interlaboratory variability using mining effluents can only be achieved through a comprehensive round-robin testing study using real-world mining effluents.

Introduction

Asking a room full of analytical chemists the question "How variable are routine chemical analyses of mine effluents and receiving waters?" is sure to trigger a lively discussion with no guarantee of a consensus emerging at the end. However, they will likely agree on one issue—variability itself is variable, and in a fairly predictable way: as analyte concentration decreases and approaches the Method Detection Limit (MDL), variability increases greatly. In fact, the MDL is a function of the variability of replicate analyses of a blank sample; and so the continuing drive toward more sensitive analytical techniques is really no more than a continuing effort to reduce analytical variability.

Aquatic toxicologists are also concerned with the variability associated with ecotoxicity tests, and have sought to minimize it by developing detailed and standardized test methods, (i.e., Environment Canada biological test methods) and the routine implementation of reference toxicant tests as tools in improving laboratory quality assurance and quality control. How variable are these tests? As with analytical chemistry, opinions vary, but a recent evaluation of a large data set involving two reference toxicants (ESG International and B. Zajdlik & Associates, 2002) indicates that aquatic toxicity tests using reference toxicants can exhibit variability in a range similar to that of routine chemical analyses (e.g., metals).

Although this is encouraging, this may not reflect reality when it comes to the variability of effluents. Analytical variability is quite variable and the MDL is important to understanding it; but toxicity testing has no obvious counterpart to the MDL (other than, possibly, the "allowable" mortality in control exposures in acute toxicity tests). So, to avoid the problem of comparing "apples and oranges", any comparison of toxicity testing variability and analytical variability needs to consider the impact of the MDL.

Another limitation in our understanding of toxicity test variability is that most of the available data involve reference toxicants, not whole effluents. As valuable as this information is,

there is no guarantee that it elucidates the variability of toxicity test data associated with mine effluents and/or receiving waters.

We informally conducted our own interlaboratory variability assessment of chemical analyses and toxicity tests, based on data available to us from several different mine operations. Our objective, using "real-world" mine effluent data, was to answer the question: is toxicity test data variability similar to analytical variability in routine analyses conducted at concentrations well above the MDL? The answer, we have found, is no; toxicity test interlaboratory variability is much higher, and is actually comparable to analytical variability at or near the MDL. This preliminary conclusion, which needs to be confirmed through a statistically rigorous study, has significant implications for the use and interpretation of toxicity test data, especially in a regulatory context.

Methods

For our assessment, we utilized existing analytical and aquatic toxicity data with mine effluents from operating mines in Canada, Peru and the United States. These data were generated over the course of the last several years in independent efforts to understand interlaboratory variability. Our principal yardstick in this assessment was the relative percent difference (RPD), commonly used by analysts for pair-wise data analysis, and defined as the difference between two values divided by the mean of the two values, expressed as a percentage. The RPD is a simple and reliable way to analyze the variability of replicate analyses.

Analytical Variability

First, we considered interlaboratory variability of chemical analyses of mine site waters by commercial analytical labs, according to widely-used and standardized analytical procedures. For the assessment, we analyzed two example data sets generated independently in Canada and Peru.

Example 1: Selenium in Canadian mine site waters.

Samples were collected at several mine sites in western Canada. The waters analyzed were a mixture of mine site and receiving waters; the analyte under investigation was selenium (Se). Data from two laboratories were compared: Lab 1 is a specialty commercial and research lab using high resolution techniques to analyze Se to a reported MDL of 10 ng/L, at a cost of US$75 per sample. Lab 2 is a commercial analytical laboratory utilizing standard analytical techniques (i.e., ICP-MS) to analyze Se with a reported MDL of 0.5 µg/L, at a cost of C$12 per analysis. The objective of the interlaboratory assessment was to determine whether the high cost of analyses conducted at Lab 1 was justified in conducting routine monitoring of Se in waters around the mine sites being studied.

In all, 30 split samples were collected from the three mine sites. Samples were collected in bottles provided by the laboratories (glass for Lab 1;HDPE for Lab 2), preserved according to their instructions (i.e., no preservation for Lab 1; nitric acid to pH 2 for Lab 2) and shipped according to standard chain-of-custody procedures. It should be noted that neither laboratory was aware that an interlaboratory comparison was being done (i.e., single-blind).

For the purposes of the assessment, "acceptable" interlaboratory variability was arbitrarily defined at an RPD of 25% or less, for concentrations in excess of 5× the MDL of Lab 2. The results are illustrated in Figures 1 and 2. The average RPD for all samples was 26.7%; however, for those samples above 2.5 µg/L, the average RPD was significantly lower (i.e., 11.4%) and for all of these pairs, the RPD was less than 25%. As Figure 3 shows, the RPD increased rapidly for those samples containing Se at concentrations below 1 µg/L. This reflects the increasing uncertainty of analyses reported by Lab 2 at concentrations approaching its MDL of 0.5 µg/L.

From these results, it was concluded that Lab 2 could be used for most routine analyses, except when receiving water data were to be interpreted strictly for conformance with the CCME (1999) aquatic life guideline of 1 µg/L, in which case the additional cost of analysis at Lab 1 would be justified in order to obtain analyses at a lower MDL.

Example 2: Peruvian mine site waters.

The second data set comprised split samples collected at three sites at a base metal mine in

north-central Peru. The water samples were analyzed for total dissolved solids (TDS), dissolved metals (Cu, Zn and Mo), total cyanide and sulfate, at commercial analytical laboratories in Canada and Peru. The objective of this interlaboratory assessment was to establish the reliability of the Peruvian laboratory relative to the Canadian laboratory, which had demonstrated stronger expertise with these analyses.

Samples were collected according to the mine's standard operating procedures and shipped to both laboratories under chain-of-custody. It was not possible to comply with all holding time requirements for the samples shipped to Canada. As in the previous example, neither laboratory was aware that they were taking part in an interlaboratory study.

A summary of the results from this example are summarized in Table 1. For this comparison, no specific criterion for acceptable RPD was established. However, adopting the arbitrary 25% limit from the previous example, we observe that out of the 18 sample pairs, five showed an RPD in excess of 25% (shown in bold in Table 1). Three of these exceedances were for dissolved zinc. What is evident is that the zinc concentrations at all three sites were low, within a factor of five of the MDL reported by Lab 2 (i.e., 0.003 mg/L). Accordingly, we expect higher variability in these analyses. For the other dissolved metals, TDS and cyanide, agreement between the two labs was excellent. It was concluded that, for these parameters, the Peruvian lab was no less reliable than the Canadian lab.

The exception was sulfate, where the RPD for two of the three analytical pairs exceeded the 25% limit (slightly). There was clearly a bias in the analyses, with Lab 1's data being consistently higher than those reported by Lab 2. The reason for the bias is unknown but could be explained by differences in the analytical methods of the two laboratories. Further analysis would have been required to determine which lab reported more accurate sulfate concentrations.

These two examples provide some indication of the variability in analytical results from different laboratories in "real-world" situations (i.e., mine site waters containing complex mixtures of cations, anions, metals and other compounds; sample collection in the field utilizing standard operating procedures; and analyses by commercial laboratories not aware that they were involved in interlaboratory assessments). Overall, interlaboratory agreement is far from perfect; but with the exception of sulfate in the second example, we know that we can be reasonably confident in such data and we can conclude that large differences among analyses of different samples reflect a difference in analyte concentrations rather than analytical variability, at least for samples that are analyzed at well above the MDL.

Table 1: Analytical data from two laboratories for analyses of mine site waters in Peru.

Analyte	Site 1			Site 2			Site 3		
	Lab 1*	Lab 2*	RPD**	Lab 1*	Lab 2*	RPD**	Lab 1*	Lab 2*	RPD**
TDS	505	466	8.0	196	172	13.0	200	178	11.6
diss. Cu	4.1	4.811	16.0	0.28	0.303	7.9	0.129	0.126	2.4
diss. Mo	0.882	0.92	4.2	0.0609	0.06	1.5	0.0327	0.03	8.6
diss. Zn	0.007	0.015	**72.7**	0.004	<0.003	**>28.6**	0.006	<0.003	**>66.7**
total CN	2.96	3.033	2.4	0.219	0.191	13.7	0.112	0.103	8.4
SO_4	275	212.4	**25.7**	33	27.1	19.6	31	23.5	**27.5**

*Concentrations expressed in mg/L
**RPD expressed as a percentage.

Aquatic Toxicity Test Variability

Now we turn our attention to the interlaboratory variability associated with aquatic toxicity tests. Again, for this assessment, we relied only upon data available to us, which came from mining operations in both the United States and Canada.

Example 3: Treated effluent, U.S. base metal mine.

This example was from a base metal mine in the western U.S. which directs all waters draining the mine site to a tailings pond. On a seasonal basis, excess water accumulating in the pond is treated by pH adjustment and removal of trace metals and solids prior to being discharged to receiving waters under the authority of an NPDES permit. The permit stipulates maximum permissible concentrations of chemical constituents and sublethal toxicity to the cladoceran, *Ceriodaphnia dubia*. The mine is required to test the treated effluent monthly when discharging water. Toxicity data are reported in toxic units (TUc), defined as an expression of the toxic strength of a substance in solution; strength being expressed as an inverse multiple of a standard endpoint of toxicity (Environment Canada, 1999); the sublethal endpoint in this case was the IC25 endpoint for *C. dubia* reproduction. Testing is conducted by commercial aquatic toxicity testing laboratories according to the standard U.S. EPA *C. dubia* test method (U.S. EPA, 1994).

After observing fluctuations in toxicity to *C. dubia* that could not easily be explained by variations in effluent chemistry, the mine started collecting split samples and conducting toxicity tests at more than one laboratory. Over the past three years, the mine consequently developed a dataset related to the interlaboratory data variability of the *C. dubia* test. Three laboratories were involved in the assessment: all samples were analyzed at Lab 1 and the splits were analyzed at either Lab 2 or Lab 3. The effluent data and, where available, the reference toxicity data reported by the labs are listed in Table 2. Data for effluents are expressed in TUc; data for the reference toxicant (i.e., sodium chloride) are expressed in mg/L.

To be consistent with our previous assessments of analytical data, we used the RPD as our measure of variability. We found the interlaboratory variability of the aquatic toxicity data with effluents to be relatively high (RPD ranged from 4.9% to 177.6%). Of the twelve data pairs, the average RPD exceeds 60%, and only one exhibited an RPD of less than 25%. This is much greater than the analytical variability presented in the previous two examples. Comparing the RPD values in Table 2 with the selenium analytical RPD values in Figure 3, we conclude that the interlaboratory variabilility of the *C. dubia* test is comparable to interlaboratory analytical variability *only* if one considers analytical variability at or near the MDL (which, as we know and have observed, is very high).

Table 2: Treated effluent and reference toxicant interlaboratory variability, U.S. base metal mine.

Effluent				Reference Toxicant (NaCl)			
Lab 1*	Lab 2*	Lab 3*	RPD***	Lab 1**	Lab 2**	Lab 3**	RPD***
4.9	6.4		25.7	578	1191		69.3
7.2	4.9		38.3	557	388		35.8
7.3	4.9		39.9	537	1203		76.6
5.5	93.4		177.6	590	N/A		N/A
7.6	5.3		36.1	524	852		47.7
5.2		4.9	4.9	427		N/A	N/A
19.3		6.7	97.4	622		670	7.4
6.1		10.9	57.0	472		910	63.4
6.1		17.2	95.2	538		800	39.2
28.4		9.8	97.4	481		640	28.4
13.0		17.0	26.5	481		640	28.4
5.7		11.8	69.7			N/A	N/A

*Effluent toxicity (IC25) reported as TUc.
**Reference toxicant IC25 data reported as mg/L..
***RPD expressed as a percentage

A review of the reference toxicant data also yields interesting results. The interlaboratory variability here (RPD ranged from 7.4% to 76.6%) certainly appears to be lower than with the effluent samples tested; but it is, overall, still higher than analytical interlaboratory variability. Is this a reflection of poor QA/QC practices in one or more of the labs conducting the test, or is it possibly a reflection of a test method with less stringent test conditions? Or, is this variability within a range acceptable for this type of test (i.e., biological vs. chemical)?

Figure 4 illustrates the data set for all 21 tests conducted at Lab 1 over the period during which the interlaboratory data were collected. The data suggest a "baseline" of between 5 and 10 TUc, with frequent excursions (i.e., samples 5, 11, 12, 16 and 19) to over 15 TUc. Are these excursions "real", that is, do they represent a change in effluent quality, or are they simply due to test variability? The fact is that we don't know. Based on the high interlaboratory variability shown in Table 2, we cannot be sure about the source of these excursions. Looking back to Figure 3, this is exactly the sort of uncertainty that one faces when trying to interpret analytical data at or near the MDL. However, when the analyte is present at concentrations substantially above the MDL, excursions such as those shown in Figure 4 are clearly "real", that is to say that they do represent a measurable change in analyte concentration.

This point can be clarified by pretending that the units of the Y-axis in Figure 4 are not TUc but µg /L, and that the data represent commercial analyses of, say, copper in receiving waters downstream from a tailings pond. Would the variations in chemistry be "real"? Yes, of course, as long as the laboratory is using suitably sensitive instrumentation (e.g., ICP-MS) and reporting analayses at a suitable MDL (which might typically be 1 µg/L), we know that they would represent measurable changes in copper concentration. We should be as confident in our toxicity data, but we aren't.

What is the significance of this assessment? Well, these toxicity data are used to monitor compliance with the terms of the mine's NPDES discharge permit. Given the high variability in the data, compliance monitoring can be an uncertain affair: you're never quite sure whether you're in or out of compliance. The implications for the mine's compliance with its discharge permit are obvious and critical.

Example 4: Treated effluent, Canadian gold mine.

Our second example of toxicity test interlaboratory variability is from a Canadian gold mine that tests its effluent using standard acute toxicity test methods for both rainbow trout and the cladoceran *D. magna* (Environment Canada, 1990a,b). The results of the paired (i.e., split samples sent to two different laboratories) tests are summarized below in Table 3[1].

Table 3: Results of rainbow trout and *D. magna* acute toxicity tests, Canadian gold mine effluent.

Organism	Mortality (%)	
	Lab 1	Lab 2
DM	0	10
RT	0	0
DM	67	100
DM	100	100
DM	46.7	100
RT	100	0
DM	46.7	30
DM	30	23.3
DM	23.3	46.7
DM	27	13

DM = *D. magna*; RT = rainbow trout.

As with the *C. dubia* data, this example data set indicates a relatively high variability between the two laboratories conducting aquatic toxicity tests. In one split effluent sample tested with rainbow trout, 100% mortality was reported by Lab 1 while zero mortality was reported by Lab 2. The apparent variability in toxicity data is much higher than analytical variability at concentrations well above the MDL, and appears to be comparable to analytical variability near the MDL. Again, consider the case where the mine needs to test its effluent for compliance with a regulatory limit. Whatever that limit may be, the variability inherent in the test may yield critical compliance issues for the mine.

[1] Here, we have avoided calculating an RPD for each data pair. The difference in between 0% and 10% mortality in a rainbow trout test is one fish. This is the same as the difference between 90% and 100% mortality; but there would be a marked difference in RPD between the two cases. Accordingly it seems appropriate not to calculate RPD.

Conclusions

Does this paper provide the final word in comparative assessments of analytical and toxicological interlaboratory variability? No. Is analytical variability always so low relative to toxicological test variability? Probably not. Is it possible that Canadian sublethal tests are less variable than their U.S. counterparts, due to more stringent test methods and respective test conditions? Yes, it is possible, as suggested in the recent TIME guidance document for acute lethality testing of mining effluents (ESG International and B. Zajdlik & Associates, 2002). Are all toxicity tests as variable as the ones we have presented? We don‘t really know (but we should know). Does this mean that toxicity testing of effluents is not warranted? No.

What this assessment suggests is that in the real world of effluent and receiving water characterization in the mining industry, toxicity tests are considerably more variable than chemical analyses used for monitoring compliance with discharge limits and water quality criteria (e.g., U.S. NPDES permits; Canadian MMERs). Accordingly, toxicity test data with mine effluents need to be interpreted with a high degree of caution, just as one must interpret analytical data for samples where the analytes are present at very low (i.e., near MDL) concentrations. The validity of any single test result is questionable.

This is an important issue with respect to compliance and it warrants further study, in view of the increasing use of aquatic toxicity testing in demonstrating compliance relating to water quality, and its use to trigger follow-up action, for instance, in the Canadian environmental effects monitoring (EEM) program. A better understanding of interlaboratory variability in aquatic toxicity tests can only be achieved through a comprehensive round-robin study using actual mining effluents. We strongly recommend the implementation of this type of study, to address the critical issues raised in this assessment.

References

CCME, 1999. Canadian Environmental Quality Guidelines. Canadian Council of Ministers of the Environment, Winnipeg, MB.

ESG International and B. Zajdlik & Associates, 2002. Guidance Document for Acute Lethality Testing of Metal Mining Effluents. Prepared for Ontario Ministry of the Environment, Toronto, ON.

Environment Canada. 1999. Guidance Document on Application and Interpretation of Single-species Tests in Environmental Toxicology. Environmental Protection, Conservation and Protection, Environment Canada, Ottawa, ON, EPS 1/RM/34.

Environment Canada. 1990a. Biological Test Method: reference method for determining acute lethality of effluents to rainbow trout. Environmental Protection, Conservation and Protection, Environment Canada, Ottawa, ON, EPS 1/RM/13 (as amended May 1996).

Environment Canada. 1990b. Biological Test Method: reference method for determining acute lethality of effluents to *Daphnia magna*. Environmental Protection, Conservation and Protection, Environment Canada, Ottawa, ON, EPS 1/RM/14 (as amended May 1996).

U.S. EPA, 1994. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms, EPA/600/4-91/002, July 1994.

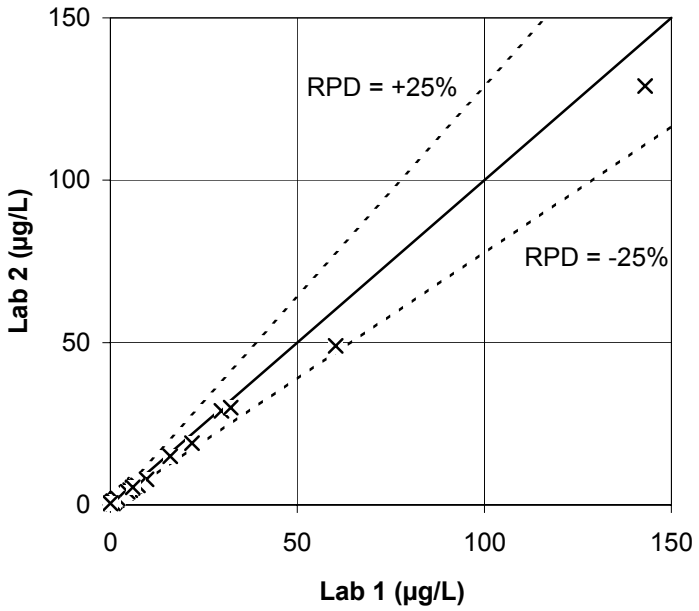

Figure 1: Se interlaboratory data for Labs 1 and 2, all data points.

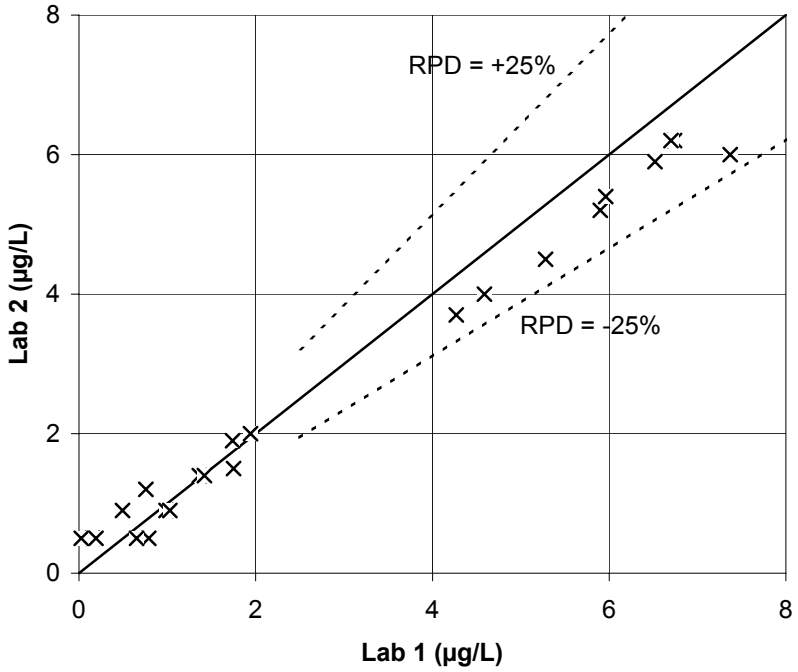

Figure 2: Se interlaboratory data for Labs 1 and 2, low concentrations only.

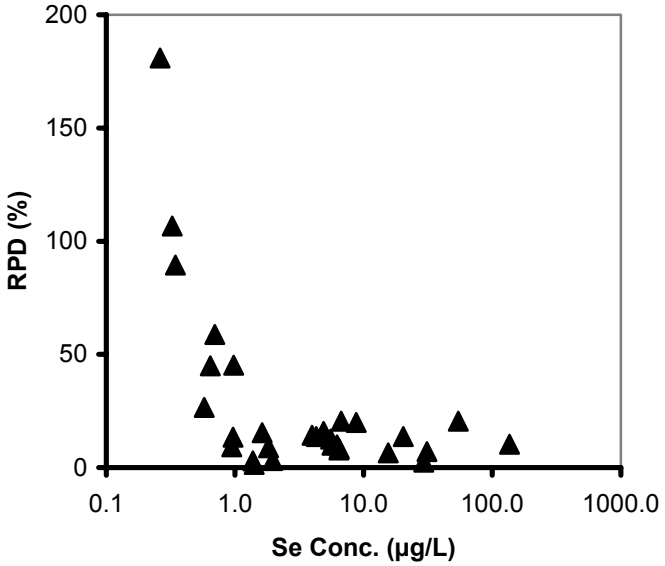

Figure 3: Interlaboratory RPD, as a function of concentration for split-sample Se analyses.

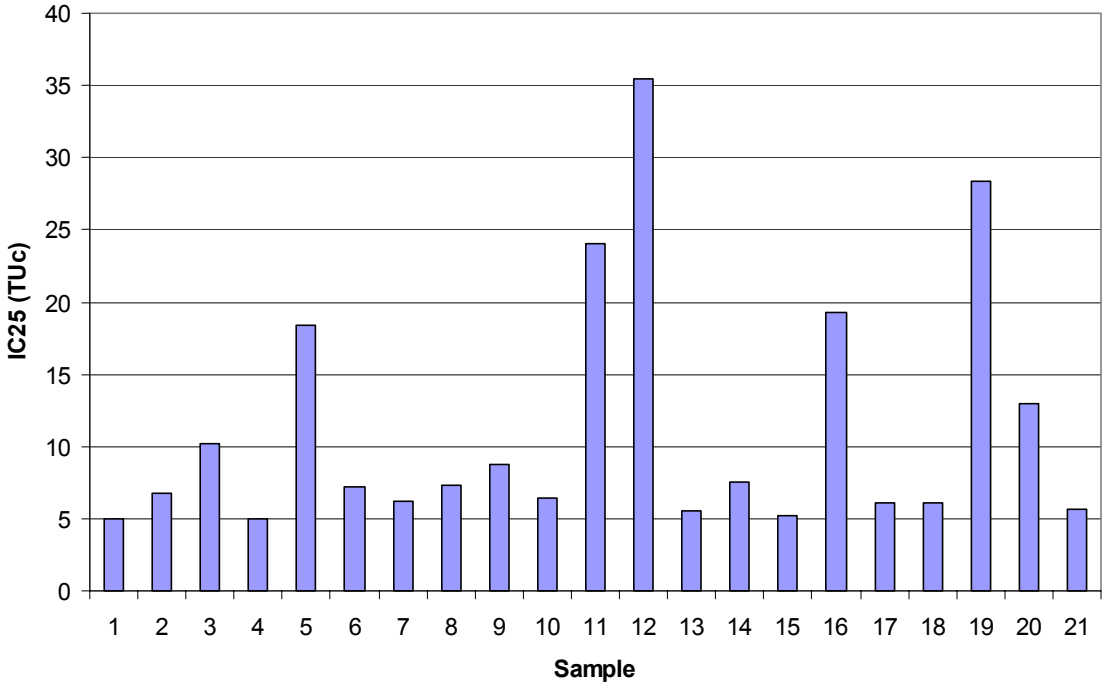

Figure 4: *C. dubia* sublethal toxicity data for Lab 1, all available data.