This page intentionally left blank

Interlaboratory comparison of sediment toxicity tests with the amphipod *Eohaustorius estuarius*

Andrew W. Jirik [1], Steven M. Bay, and Stan Asato[2]

ABSTRACT

An interlaboratory comparison exercise was conducted among seven laboratories in order to document the reproducibility of sediment toxicity measurements conducted during the Bight'98 regional sediment survey. Sediments from four stations in Los Angeles/Long Beach Harbor were tested using a 10-d survival test of the amphipod *Eohaustorius estuarius*. All laboratories successfully performed the sediment test and associated reference toxicant test. While statistically significant differences were found in amphipod mean survival rates among some laboratories for the field-collected sediments, no consistent significant bias was observed. Testing by multiple laboratories did not appear to reduce the precision of the results. The laboratories demonstrated excellent concordance (Kendall's W = 0.91) in ranking the field-collected sediments by toxicity. Agreement on classifying the sediments into categories (nontoxic, moderately toxic, and highly toxic) based upon the percent of survival was best for highly toxic sediments. An analysis of test precision based upon the variance among replicates within a test indicated that the measured survival rate for a sample may vary by up to 12 percentage points from the actual response.

INTRODUCTION

Laboratory tests that measure the toxic effects of sediments on benthic organisms are being used with increasing frequency to help assess the impact of contaminated sediments on marine life. The most frequently used test measures the survival rate of amphipods following a 10-d sediment exposure. Standardized methods for conducting amphipod tests have been published (U.S. EPA 1994), and these tests are integral components of national programs to assess sediment quality in coastal areas of the United States (Long 2000). Amphipod toxicity tests have been used in southern California to examine temporal changes in sediment quality associated with pollution reductions (Swartz *et al.* 1986, Bay 1992), identify toxic hot spots (Fairey *et al.* 1996), and estimate the spatial extent of sediment toxicity (Anderson *et al.* 1998, Bay *et al.* 1998). Amphipod toxicity tests were an integral part of the Bight'98 regional survey, a cooperative regional survey that included the testing of 241 sediment samples by 7 laboratories (Bight'98 Steering Committee 1998).

Toxicity tests using biological responses to measure effects provide valuable information about the significance of chemical contamination. The response of an organism to contaminated sediment provides an integrated measure of effect that reflects the combined action of all materials (measured and unmeasured) present and also takes into consideration site-specific variations that affect toxicity of contaminants (e.g., binding to particles, changes in chemical speciation). However, the use of an organism as the "detector" creates the potential for increased uncertainty in the results. In addition to the sources of variability present in most laboratory procedures (e.g., measurement error, variations in sample composition, or preparation), variability may be produced due to uncontrolled variations in the sensitivity of the test organism. Additional sources of variability may be introduced when data from multiple laborato-

This picture not available at this time

[1]*Present address: Port of Los Angeles, 425 S. Palos Verdes Street San Pedro, CA 97031*

[2]*City of Los Angeles, Hyperion Treatment Facility, 12000 Vista del Mar, Playa del Rey, CA 90293*

ries are combined for use in cooperative monitoring programs or nationwide assessments of sediment quality data obtained from disparate sources (e.g., Long *et al.* 1995). Thus, the reproducibility of sediment toxicity tests is an area of concern.

The results of interlaboratory comparison studies to examine the reproducibility of marine amphipod toxicity tests have been reported twice (Mearns *et al.* 1986, Schlekat *et al.* 1995). These studies included participants having extensive experience with the test species and they demonstrated that amphipod toxicity tests are reliable and reproducible methods. However, these studies do not provide sufficient information to address questions about the comparability of tests conducted by laboratories with variable levels of experience, a situation typical of monitoring programs. For example, while all seven laboratories conducting amphipod toxicity tests for Bight'98 were experienced in toxicity test procedures, only one had previously worked with the specific test species used (*Eohaustorius estuarius*), and two of the laboratories had not previously conducted sediment toxicity tests with amphipods. Inexperience with the toxicity test method or the species may influence the test results through the inappropriate acclimation or handling of the test organisms or errors in measurement of the test endpoint (e.g., amphipod survival).

Information describing interlaboratory comparisons with *E. estuarius* is scarce; only one interlaboratory comparison study using this species has been published. That study tested dilutions of a single highly toxic stored field sediment. Consequently, no information is available that describes the reproducibility of *E. estuarius* test results using field sediments that are representative of those tested in most monitoring programs and regional surveys.

This article presents the results of the interlaboratory comparison exercise conducted as part of the Bight'98 regional survey. The study had three objectives. The first objective was to assess whether each laboratory was able to perform the 10-d *Eohaustorius* test in accordance with pre-established standards. This objective was evaluated by examining the attainment of test acceptability criteria for reference toxicant and sediment test procedures. The second objective was to assess the degree of agreement among laboratories for sediment toxicity results under conditions typical of a regional survey. This objective was accomplished by comparing the toxicity results among laboratories for a reference toxicant and also for field sediments that were collected using standardized methods. The third objective was to evaluate whether the participation of multiple laboratories introduced greater variability in test results compared to analyses by a single laboratory. This issue was investigated by comparing variability in toxicity test responses to field sediments and a reference material within and among laboratories.

METHODS

Experimental Design

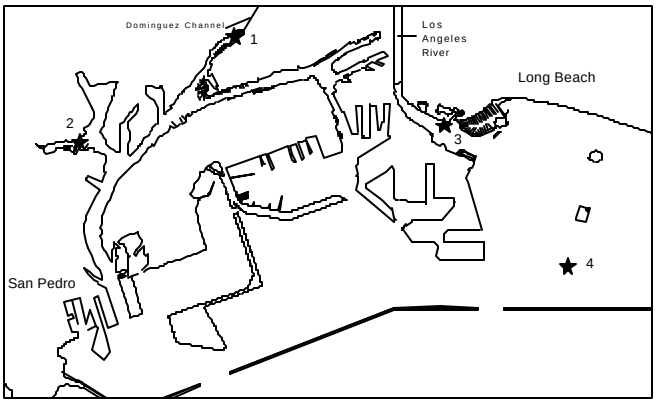
Seven laboratories participated in the interlaboratory study: Aquatic Bioassay and Consulting Labs (Ventura, CA), Marine Pollution Studies Laboratory (University of California), City of Los Angeles Environmental Monitoring Division, City of San Diego Ocean Monitoring Program, MEC Analytical Systems (Carlsbad, CA), Orange County Sanitation District, and SCCWRP. Sediment samples from four locations in Los Angeles/Long Beach Harbor were selected based upon prior data (SWRCB 1996) to represent a gradient of toxicity ranging from highly toxic (<50% survival) to moderately toxic (approximately 70% survival). Several steps were taken to minimize sources of variation not associated with the laboratories: all sediment collection and processing was conducted by one laboratory, test organisms were provided by a common supplier, tests were conducted at the same time, and reference toxicant solutions were prepared and distributed from a central source.

Sample Collection and Handling

Sediment samples were collected on May 7, 1998, from four stations: LAH1 (33° 46.564′, 118° 14.608′), LAH2 (33° 45.343′, 118° 16.787′), LAH3 (33° 45.529′, 118° 11.684′), and LAH4 (33° 43.890′ 118° 9.955′) (Figure 1). Sediment samples were collected with a 0.1 m^2 modified van Veen grab. At a given station, multiple grab samples were taken to provide 12 L of sediment. A plastic (high density polyethylene) scoop was used to collect sediment from the top 4 cm of the undisturbed surface material in the grab. The sediment was transported to the laboratory in polycarbonate containers on ice. Once back at the laboratory, the sediment was homogenized in a polyethylene bucket using an overhead mixer. The homogenized sediment was transferred to polyethylene containers. Each laboratory received 1.2 L of sediment from each of the four stations.

Test organisms, *Eohaustorius estuarius*, were obtained from Northwestern Aquatic Sciences (collection site: Beaver Creek, Oregon). The animals and collection site sediment were collected on May 6 and shipped by overnight courier to each laboratory. The organisms were transferred into aquaria at each laboratory and acclimated under conditions of 20 ppt salinity, aeration, constant illumination, and 15° C until the initiation of the test.

FIGURE 1. Location of sites sampled in Los Angeles/ Long Beach Harbor for interlaboratory toxicity comparison.



Test Procedures

Sediment toxicity was determined using a 10-d amphipod survival test (U.S. EPA 1994). Sediment toxicity tests were conducted in 1 L glass test containers. Sediment was added to the test containers 1 d prior to the start of the test (May 11, 1998). Sediment samples were mixed thoroughly and then added to the test containers to form a sediment layer approximately 2 cm deep. Filtered seawater (20 ppt) was added slowly until a final volume of 800 mL was reached. Pipettes connected to an air source provided aeration. Test containers were then allowed to equilibrate overnight. Each sample consisted of five randomly arranged replicates, along with an extra container to provide samples for water quality. A negative control (consisting of test animal collection site sediment) was included in each batch of samples tested.

At the start of the test (May 12, 1998), 20 amphipods were added randomly to each test container. Test animals were exposed to the sediment samples for 10 d at 15° C under constant illumination. Test containers were checked daily for air and for any dead animals or animals stuck to the surface of the water. Any floating animals were submerged by gently pushing them beneath the surface with a probe. At the end of the exposure period (May 22, 1998) the sediment was screened through a 0.5 mm screen and the number of surviving amphipods was recorded.

Concurrently with the sediment toxicity test, a cadmium reference toxicity test was conducted. The aqueous phase reference toxicant test consisted of three replicates of five dilutions, plus the control sample. The concentrations were 0.32, 1.00, 3.20, 5.60, and 10.00 mg/L. A sample of the 10.00 mg/L concentration was analyzed for verification purposes. At the beginning of the test (May 12, 1998), 10 amphipods per replicate were added randomly to each test container and exposed to the reference toxicant for 4 d. At the end of 4 d (May 16, 1998), the number of surviving animals was recorded and the LC_{50} (median lethal concentration) was calculated.

Initial water quality (temperature, pH, dissolved oxygen, salinity, and total ammonia) was measured on samples of overlying water and interstitial water from the extra test container. Interstitial water samples were obtained by centrifuging a sediment sample at 3,000 X g for 20 min. Temperature, pH, salinity, and dissolved oxygen of the overlying water were also recorded at the end of the exposure period. All water quality indicators were measured with laboratory-approved equipment and procedures. Water quality measurements for the reference toxicant test were similar to those used in the sediment phase of the test.

Data Analysis

The mean percent of survival was calculated for the five replicates tested for each sediment sample. The data were then normalized by dividing the sample mean rate by the appropriate control mean rate. This value, expressed as a percentage of the control response, reduces the variation in results due to differences in control survival and facilitates comparisons among different tests. T-tests were conducted versus the negative control to determine significance at the 95% level.

Mean survival was also calculated for each cadmium concentration of the reference toxicant test. Reference toxicant test LC_{50} values were calculated with the Spearman-Karber method and compared among laboratories. An individual test result within two standard deviations of the mean for all laboratories was considered acceptable.

Sediment toxicity was defined by two criteria: (1) a statistically significant difference between the sample and control and (2) a minimum percentage difference between the sample and control. Samples that were significantly different from the control and had a 20% response relative to the control (survival rate less than 80% of the control) were classified as toxic. This measure of toxicity represents a 90% power to determine statistical significance in survival between control and sample (Thursby *et al.* 1997). Toxic samples were further classified as moderately toxic (50 to 79% survival) and highly toxic (less than 50% survival).

The agreement among the laboratories for individual sediments was first assessed by analysis of variance (ANOVA) for data meeting assumptions of normality, otherwise by a Kruskal-Wallis ANOVA on ranks. In cases where significant differences were found, Tukey pairwise multiple comparison tests were conducted to detect specific differences between the laboratories.

The degree of association of toxicity rankings among laboratories was assessed by Kendall's coefficient of

concordance (W). The field sediments were ranked in order of toxicity for each laboratory, with a value of 1.0 assigned to the sediment with the highest survival rate and a value of 4.0 assigned to the sediment with the lowest survival rate. Kendall's W ranges from 0.0 (no degree of association) to 1.0 (perfect association).

Variability in field sediment results among the laboratories was assessed by calculating the pooled variance of the mean percent of survival for each station. This variance was compared to the variability within a laboratory, calculated as the pooled variance among replicates for each laboratory.

Results

Test Success

All of the participating laboratories met the test acceptability criteria for the sediment and reference toxicant tests. Each laboratory obtained nearly 100% survival in the collection site control sediment and recorded at least 90% survival for the reference toxicant seawater control (Table 1). The laboratories reported that all of the experiments were conducted within the parameters of the test protocol (including water quality) and that no tests had to be repeated.

Interlaboratory Comparability

Reference Toxicant

Results of the reference toxicant tests varied approximately five-fold among laboratories. The LC_{50} values ranged from 1.8 to 9.43 µg/L; five of the seven laboratories, however, had LC_{50} values that were close (±28%) to the mean (5.39 µg/L) (Table 2). The most sensitive test result (lowest LC_{50}) was reported by Lab 1; its LC_{50} was significantly less than the values reported by the other laboratories. The 95% confidence limits for the results from the other laboratories overlapped one another, indicating that the results were similar. None of the results were classified as an outlier, since all were within two standard deviations of the mean value.

TABLE 1. Laboratory control performance.

Lab	Sediment Control Survival (%)	Reference Toxicant Control Survival (%)
1	99	100
2	100	100
3	98	97
4	96	90
5	97	100
6	100	100
7	97	97

Sediments

Each of the four sediments was consistently identified as toxic by the laboratories. A statistically significant difference in the amphipod survival rate relative to the control sample (t-test, $p<0.05$) was obtained for all but 1 of the 28 samples evaluated in the study (4 sediments times 7 laboratories).

TABLE 2. Cadmium reference toxicant results obtained during the interlaboratory comparison exercise.

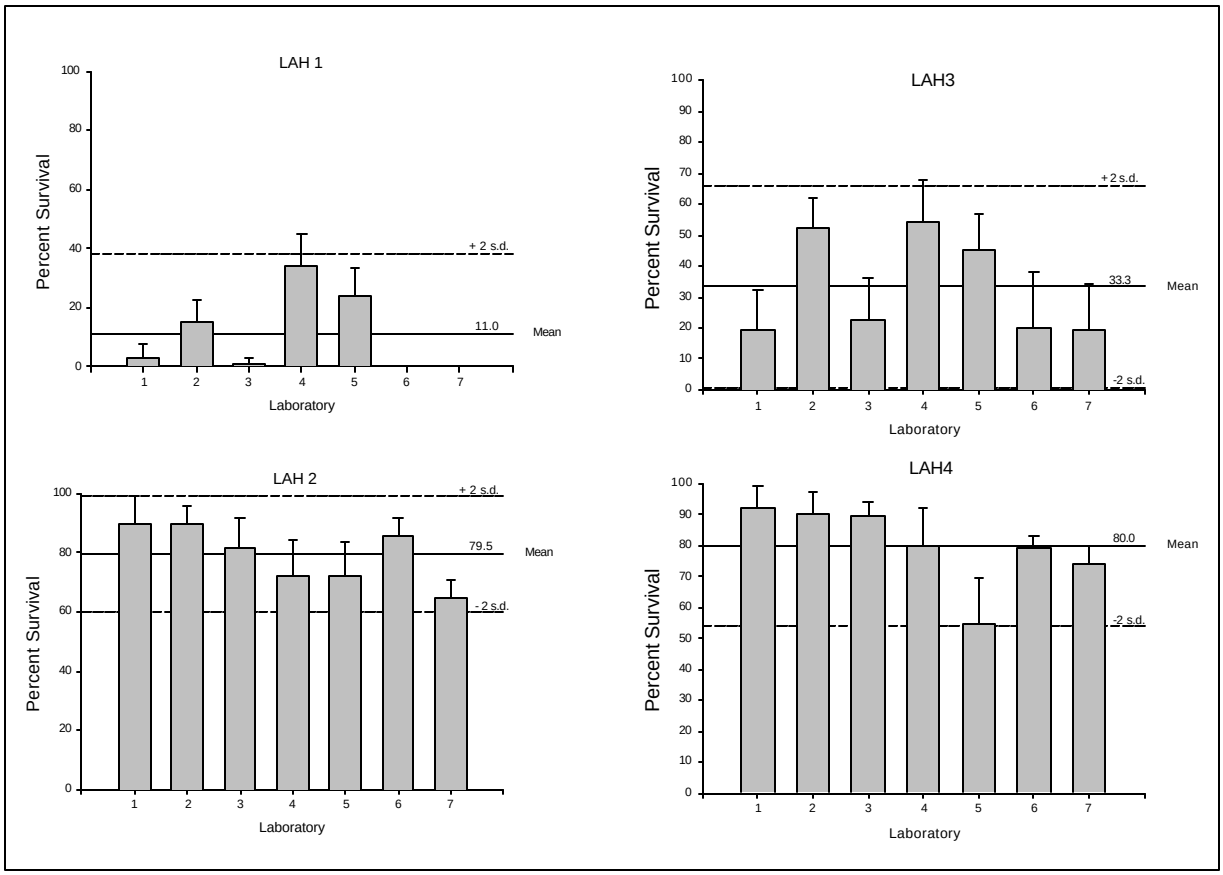
Lab	LC50 (mg/L)	95% CI
1	1.8	1.6 - 1.9
2	9.4	7.1 - 12.6
3	4	3.2 - 5.0
4	5.7	4.6 - 7.0
5	5.9	5.0 - 7.0
6	6.3	4.6 - 8.6
7	4.7	3.9 - 5.7

A range of toxicity was present among sediment types. High mortality (<35% mean survival among laboratories) was produced by sediments from two stations, LAH1 and LAH3 (Figure 2). Sediment from LAH1 was most toxic; a mean survival rate of 11% was obtained for this sample and four laboratories reported <5% survival. Low-moderate toxicity was detected in the remaining stations (LAH2 and LAH4). Most laboratories reported 70-90% survival in tests with these samples (Figure 2), and a mean survival of 80% was obtained among laboratories for each station.

A similar amount of variation among laboratories was present for each sample. Standard deviations for stations LAH1, LAH2, LAH3, and LAH4 were 13.7, 9.9, 16.4, and 13.0, respectively. An overall estimate of the variability of the toxicity results among laboratories was calculated as the grand mean of the variances for all 28 toxicity measurements (4 field sediments times 7 laboratories), which was 93.9. The grand mean variance was used to calculate a 95% confidence interval for amphipod survival of ±12.0%. Thus, the percentage of survival measured by a laboratory is expected to be within 12 percentage points of the true value 95% of the time. The coefficients of variance (CVs) among laboratories (standard deviation expressed as a percentage of the mean) were dependent upon the station (12.4-163%), with the highest values corresponding to the stations with the lowest mean rates of survival.

The individual sample results among laboratories were compared using three approaches, each reflecting a different method of data interpretation. For the first approach, the percent of survival results among laboratories were compared. Examination of the survival results indicated that some bias may have been present in the data from two laboratories. Lab 7 reported the lowest or second lowest survival rate for each of the sediment samples and Lab 2 reported the highest or second highest survival rate for three samples (Table 3). These differences were not statistically significant, however. Analysis of variance followed by Tukey pairwise comparisons indicated that the

FIGURE 2. Survival results for *Eohaustorius estuarius* exposed to field sediments (LAH1-LAH4). Bars represent the mean of five replicates tested at each laboratory. Error bars are one standard deviation. Solid reference line is the consensus mean of the seven laboratories. Dotted reference lines are ±2 standard deviations from the consensus mean.



results for Labs 2 and 7 were similar to those for most of the other laboratories (Table 3). Significant differences in the percent of survival were detected between laboratories in 25% of the 84 possible pairwise comparisons, but no consistent pattern was observed among laboratories in these differences. Only one case (Lab 5 for Station LAH4) was encountered where the results for a laboratory were different from all other laboratories (Table 3).

TABLE 3. Laboratories arranged by order of survival results for each field sediment. Laboratories not significantly different from one another (Tukey pairwise comparison, p>0.05) are connected by solid lines.

Sediment	Laboratory number						
	Highest Survival						Lowest Survival
LAH1	4	5	2	1	3	6	7
LAH2	2	1	6	3	5	4	7
LAH3	4	2	5	3	6	7	1
LAH4	1	2	3	4	6	7	5

The second evaluation approach examined the ability of the laboratories to assess the relative toxicity of the four samples. The sediments were assigned ranks based upon the percent of survival results. Each of the seven laboratories ranked LAH1 and LAH3 as the most toxic and next most toxic samples, respectively (Table 4). Four of the laboratories ranked the sediments in exactly the same order and a fifth only differed in that LAH2 and LAH4 were tied in the rankings. The Kendall coefficient of concordance based upon these data was 0.91, indicating a high level of agreement ($p<0.01$) among laboratories.

The final assessment approach examined the ability of the laboratories to classify the degree of sediment toxicity using response thresholds (non-toxic, moderately toxic, and highly toxic) typical of regional monitoring programs. The agreement among laboratories in classifying the samples

TABLE 4. Rank of field sediments by amphipod survival. 1 = highest survival, 4 = lowest survival.

	Laboratory Number						
Sediment	1	2	3	4	5	6	7
LAH1	4	4	4	4	4	4	4
LAH2	2	1.5	2	2	1	1	2
LAH3	3	3	3	3	3	3	3
LAH4	1	1.5	1	1	2	2	1

varied and was dependent upon the magnitude of toxicity. All seven laboratories classified sediment from LAH1 (the most toxic sediment) as highly toxic (Table 5). Relatively consistent results were also obtained for Station LAH3; all seven laboratories classified LAH3 as toxic, with five classifying it as highly toxic and two as moderately toxic. The classification results were more variable for LAH2 and LAH4; approximately half of the laboratories placed these sediments into the non-toxic category (Table 5). The mean survival rate (among laboratories) for these two stations was 80%, the same as the response threshold used to distinguish between non-toxic and moderately toxic samples.

Interlaboratory Variability

The reference toxicant test results reported by the interlaboratory study participants were similar to the results of multiple tests conducted by a single laboratory (Lab 6), as shown in Figure 3. Both the intralaboratory and interlaboratory data sets had similar means (5.4 and 6.4, respectively) and ranges (1.8-9.4 and 2.2-10, respectively), indicating that the participation of multiple laboratories had little effect on the reference toxicant results.

The assessment of the amount of variation in the data attributable to interlaboratory variability was complicated by the lack of repeated analyses of the same samples by a single laboratory. As an alternative, the pooled variance among replicates within each laboratory was compared to the pooled variance in mean survival among laboratories for each station. The interlaboratory variance was higher than the replicate variance for each of the four sediment samples (Figure 4). Station LAH1 (the most toxic sample) showed the greatest difference, with the interlaboratory variance being 4.2 times greater than the within-replicate variance. The interlaboratory variance was 1.4 to 2.2 times greater than the replicate variance for the other three stations.

TABLE 5. Comparison of individual laboratory classifications of samples tested. N=nontoxic (³80% survival), MT=moderately toxic (50-79% survival, significantly different from the control (t-test, p<0.05)), HT=highly toxic (<50% survival, significantly different from the control (t-test, p<0.05)).

	Sample			
Lab	LAH1	LAH2	LAH3	LAH4
1	HT	N	HT	N
2	HT	N	MT	N
3	HT	N	HT	N
4	HT	MT	MT	MT
5	HT	MT	HT	MT
6	HT	N	HT	MT
7	HT	MT	HT	MT

DISCUSSION

This study has shown that the *E. estuarius* amphipod toxicity test can be conducted with a high degree of success and reproducibility, even among laboratories with varying

FIGURE 3. Comparison of *Eohaustorius estuarius* 96-h Cadmium LC_{50} data for multiple tests conducted at one laboratory (open symbols) to data from the Bight'98 interlaboratory comparison (filled circles).

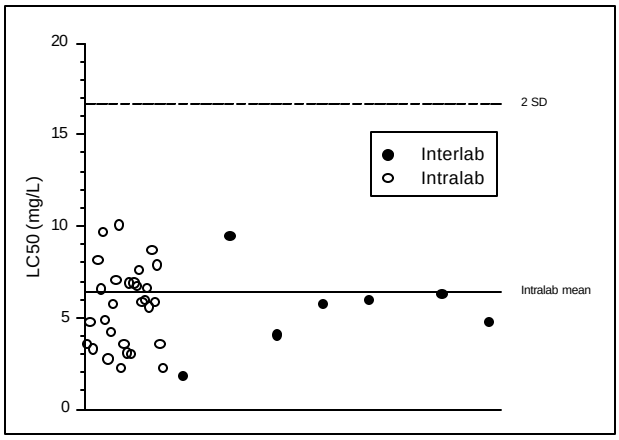
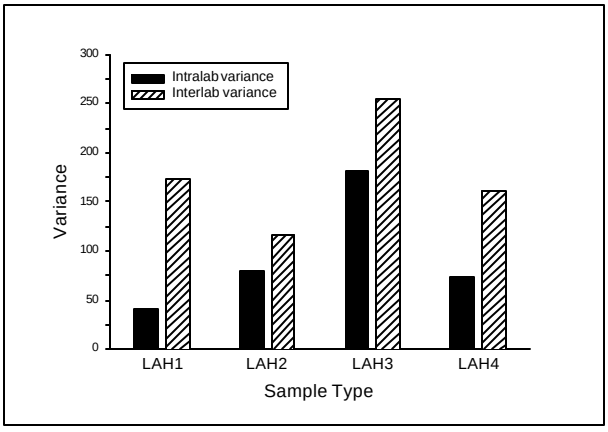


FIGURE 4. Comparison of intralaboratory and interlaboratory variance for sediment toxicity results. The values represent the pooled variance of the percent of survival data either among replicates within a laboratory (intralaboratory) or among means from the seven participating laboratories (interlaboratory).



levels of experience with the method. Each laboratory met the performance criteria specified by the protocol (U.S. EPA 1994) and was able to discriminate statistically between the control sediment and toxic field-collected sediments.

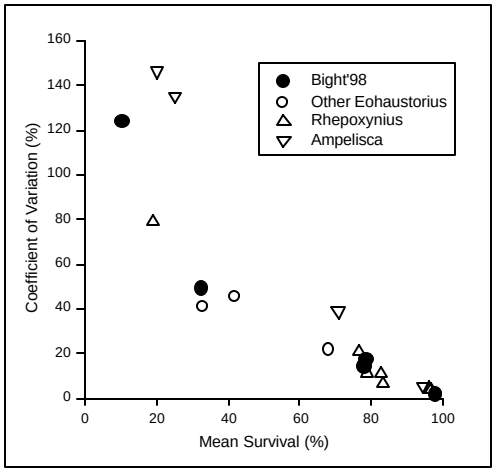
Variable results were obtained among some of the laboratories, as shown by statistically significant differences in the results (percent of amphipod survival) for the same sediment sample. However, the variability measured in this study is similar to that found in other interlaboratory studies. The CVs reported here are very similar to those reported in interlaboratory comparisons using other species of amphipods (Figure 5, Mearns *et al.* 1986, Schlekat *et al.* 1995). All of these studies show that the relative variability (CV) increases markedly when toxicity is high. Conversely, interlaboratory variability is least in low-moderate toxicity samples, which constitute the bulk of sediments present in coastal waters (Long *et al.* 1996).

No specific cause was identified for the variability in sediment test results observed among laboratories. The lack of a strong bias in the interlaboratory results, indicated by the similarity of station rankings (Table 4) and the similarity of intralaboratory and interlaboratory reference toxicant results, indicates that laboratory-specific differences in organism sensitivity or test methods were not a major factor. The variability may have been due to factors such as changes in contaminant bioavailability due to sediment storage and handling or may reflect the inherent variability of the test organism response.

While these results indicate that the precision within one test is greater than the precision among multiple tests conducted by different laboratories, the data cannot be used to determine whether this variability is introduced by multiple tests or by multiple laboratories. Multiple tests within one laboratory may produce a similar level of variability. A measurement of intralaboratory variability, obtained through the repeated measurement of the same sample, is needed to provide a comparison with the interlaboratory variability measured in this study. Such a study is difficult to conduct with sediments because toxicity may be altered by long-term sediment storage.

A degree of uncertainty must be expected for any study that employs tests that have variability (Mearns *et al.* 1986). The variability of toxicity tests is similar to that of chemical analysis, indicating that both types of measurements have similar reliability. The significance of this variability depends on the way in which the data are interpreted. If the results are used to rank or otherwise describe the relative toxicity of multiple stations, then there is good agreement among laboratories, especially regarding the identification of the most toxic stations ("hot spots"). For regional assessments, thresholds based upon percent of survival are often used to identify one or more levels of toxic response (Bay *et al.* 1998, Fairey *et al.* 1996, Long *et al.* 1998). Our data, representing contaminated field sediments spanning the range of toxicity typically encountered, indicate that survival measurements are likely to vary by up to 12 percentage points from the actual value. This variability may alter the classification of a sediment sample whose true level of toxicity is near the threshold value. This uncertainty is not a unique problem to toxicity tests and can be minimized by stratified sampling designs that utilize the information from multiple sediment samples to characterize the extent and magnitude of toxicity within a region.

FIGURE 5. Interlaboratory variability as a function of mean survival for three commonly applied 10-d amphipod tests. In addition to the current study, data are from Mearns *et al.* (1986) and Schlekat *et al.* (1995).



LITERATURE CITED

Anderson, B.S., J.W. Hunt, S. Tudor, J. Newman, R.S. Tjeerdema, R.S. Fairey, J. Oakden, C. Bretz, C.J. Wilson, F. LaCaro, M.D. Stephenson, H.M. Puckett, E.R. Long, T. Fleming and K. Summers. 1998. Chemistry, Toxicity, and Benthic Community Conditions in Selected Sediments of the Southern California Bays and Estuaries. Final Report. State Water Resources Control Board. Sacramento, CA.

Bay, S. 1992. Temporal and spatial changes in sediment toxicity in Santa Monica Bay. pp. 81-87 *in:* J. Cross (ed.), Southern California Coastal Water Research Project. Annual Report 1990-1991 and 1991-1992. Westminster, CA.

Bay, S.M., D.J. Greenstein, A.W. Jirik and J.S. Brown. 1998. Southern California Bight 1994 Pilot Project: VI. Sediment toxicity. Southern California Coastal Water Research Project. Westminster, CA.

Bight '98 Steering Committee. 1998. Southern California Bight 1998 Regional Monitoring Survey (Bight '98): Coastal ecology workplan. Southern California Coastal Water Research Project. Westminster, CA.

California State Water Resources Control Board (SWRCB). 1996. Sediment Chemistry, Toxicity, and Benthic Community Conditions in Selected Water Bodies of the Los Angeles Region. Final Report. Sacramento, CA.

Fairey, R.S., C. Bretz, S. Lamerdin, J.W. Hunt, B.S. Anderson, S. Tudor, C.J. Wilson, F. LaCaro, M.D. Stephenson, H.M. Puckett and E.R. Long. 1996. Chemistry, Toxicity, and Benthic Community Conditions of Selected Sediments of the San Diego Bay Region. Final Report. State Water Resources Control Board. Sacramento, CA.

Long, E.R. 2000. Degraded sediment quality in U.S. estuaries: A review of magnitude and ecological implications. *Ecological Applications* 10:338-349.

Long, E.R., L.J. Field and D.D. MacDonald. 1998. Predicting toxicity in marine sediments with numerical sediment quality guidelines. *Environmental Toxicology and Chemistry* 17:714-727.

Long, E.R., D. MacDonald, S. Smith and F. Calder. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environmental Management* 19:81-97.

Long, E.R., A. Robertson, D. Wolfe, J. Haeedi and G. Sloane. 1996. Estimates of the spatial extent of sediment toxicity in major U.S. estuaries. *Environmental Science and Technology* 30:3585-3592.

Mearns, A.J., R.C. Swartz, J.M. Cummins, P.A. Dinnel, P. Plesha and P.M. Chapman. 1986. Inter-laboratory comparison of a sediment toxicity test using the marine amphipod *Rhepoxynius abronius*. *Marine Environmental Research* 19:13-37.

Schlekat, C.E., K.J. Scott, R.C. Swartz, B. Albrecht, L. Antrim, K. Doe, S. Douglas, J.A. Ferretti, D.J. Hansen, D.W. Moore, C. Mueller and A. Tang. 1995. Interlaboratory comparison of a 10-day sediment toxicity test method using *Ampelisca abdita*, *Eohaustorius estuarius*, and *Leptocheirus plumulosus*. *Environmental Toxicology and Chemistry* 14:2163-2174.

Swartz, R.C., F.A. Cole, D.W. Schultz and W.A. DeBen. 1986. Ecological changes in the Southern California Bight near a large sewage outfall: Benthic conditions in 1980 and 1983. *Marine Ecology Progress Series* 31:1-13.

Thursby, G.B., J. Heltshe and K.J. Scott. 1997. Revised approach to toxicity test acceptability criteria using a statistical performance assessment. *Environmental Toxicology and Chemistry* 16:1322-1329.

United States Environmental Protection Agency (U.S. EPA). Methods for Assessing the Toxicity of Sediment-associated Contaminants with Estuarine and Marine Amphipods. 1994. EPA/600/R-94/025. Washington, DC.

ACKNOWLEDGEMENTS

The authors wish to thank the laboratory personnel of the following organizations who conducted the bioassays for this study: Aquatic Bioassay and Consulting Laboratory (ABCL), Marine Pollution Studies Laboratory (MPSL), City of Los Angeles Environmental Monitoring Division, City of San Diego Ocean Monitoring Program, MEC Analytical Systems, Orange County Sanitation District, and SCCWRP. We thank Darrin Greenstein (SCCWRP) for coordinating the logistics of the exercise and Tim Mikel (ABCL) for assisting with the reference toxicant data analysis. We thank Molly Leecaster (SCCWRP) for her valuable input regarding statistical analysis. Dario Diehl (SCCWRP) as well as the skipper and crew of the *Ocean Sentinel* (County Sanitation Districts of Los Angeles County) ably assisted in the collection of field sediments. We are also grateful to Brian Anderson (MPSL) for sharing his insights into working with *Eohaustorius estuarius* as a test species.