SETAC PRESS

Environmental Toxicology and Chemistry, Vol. 19, No. 10, pp. 2598–2601, 2000
Printed in the USA
0730-7268/00 $9.00 + .00

CLASSIFYING PROBABILITIES OF ACUTE TOXICITY IN MARINE SEDIMENTS WITH EMPIRICALLY DERIVED SEDIMENT QUALITY GUIDELINES

EDWARD R. LONG,*† DONALD D. MACDONALD,‡ CORINNE G. SEVERN,§ and CAROLYN B. HONG§
†National Oceanic and Atmospheric Administration, National Centers for Coastal Ocean Science, 7600 Sand Point Way NE, Seattle, Washington 98115, USA
‡MacDonald Environmental Sciences Ltd., 2376 Yellow Point Road, R.R. 3, Nanaimo, British Columbia, Canada V9X 1W5
§EVS Environment Consultants, 200 West Mercer Street, Suite 403 Seattle, Washington 98119, USA

(*Received* 15 *June* 1999; *Accepted* 18 *April* 2000)

Abstract—Matching, marine sediment chemistry, and toxicity data (n = 1,513), compiled from three studies conducted in the United States, were analyzed to determine both the frequency of acute toxicity to amphipods and average percentage survival in laboratory bioassays within ranges in toxicant concentrations. We determined that the probability of observing acute toxicity was relatively low (<10%) and that average control-adjusted survival equaled or exceeded 92% in samples in which sediment quality guidelines were not exceeded. Both the incidence of toxicity increased and average survival decreased as chemical concentrations increased relative to the guidelines. In sediments with highest contaminant concentrations, 73 to 83% of the samples were highly toxic, and average control-adjusted amphipod survival was 37 to 46%. Results of this study confirm that the relationships between sediment chemical concentrations and toxicity reported in a previous study were robust. Further, they indicate that numerical guidelines for saltwater sediments can be used to estimate the probability of observing toxic effects in acute amphipod tests.

Keywords—Sediment quality guidelines Sediment toxicity

INTRODUCTION

Empirically derived sediment quality guidelines (SQGs) were developed on the basis of the associations observed between measures of adverse biological effects and the concentrations of potentially toxic substances in marine and estuarine sediments. Two sets of SQGs developed for saltwater include the effects range—low (ERL) and effects range—median (ERM) values for 25 chemicals [1] and the threshold effects level (TEL) and probable effects level (PEL) values for 31 chemicals [2]. The relative abilities of individual SQGs to identify chemical concentrations that were either rarely or frequently associated with adverse effects were previously reported [1,2], using the data with which the guidelines were derived. Subsequently, the predictive abilities of the SQGs were quantified [3] using an independent data set (i.e., not used in the guidelines derivation). Data used to calculate the predictive abilities of the SQGs were compiled in a large database from studies conducted in estuaries of the Atlantic, Gulf of Mexico, and Pacific coasts. All these studies provided matching data on chemical contamination and acute toxicity using comparable methods (n = 1,068).

Recommended applications of the SQGs were based on observations of their predictive abilities [4]. In that study, four independent indices of chemical contamination were identified that defined four categorical ranges in the incidence of acute toxicity in amphipod survival tests. Acute toxicity measured in laboratory bioassays occurred rarely (10–12% of samples) in sediments in which none of the guidelines was exceeded. The percentages of samples in which acute toxicity was observed increased with the numbers of guidelines exceeded. In addition, they increased with the concentrations of mixtures of substances normalized to the guidelines (expressed as mean SQG quotients). This classification system provided users of the guidelines with a basis for estimating the probabilities that sediment samples would be acutely toxic in laboratory tests using the data on chemical concentrations alone.

The primary objective of this paper was to further evaluate the predictive abilities of the SQGs using an expanded database and therefore to determine the robustness of the chemistry/toxicity relationships observed in the previous study. Another objective was to summarize the probabilities of observing acute toxicity to amphipods on the basis of evaluations of the sediment chemistry data, using the guidelines. Finally, to determine the applicability of the guidelines in sediments with somewhat atypical geochemical properties, we analyzed data available from Biscayne Bay, Florida, USA, and Pearl Harbor, Hawaii, USA.

METHODS

Data from solid-phase tests of sediments were selected for these evaluations because of their relatively low uncertainty and many other favorable attributes [5]. The most data available nationwide are from tests performed with the amphipod *Ampelisca abdita.* A database of matching sediment chemistry and toxicity data was compiled from three sources. The majority of the data (consisting of information from 1,068 samples collected in estuaries of the Atlantic, eastern United States; Gulf of Mexico, southern United States; and Pacific coast, western United States) was compiled for a previous analysis of the predictive ability of sediment guidelines [3]. These data are hereafter referred to as the national data set.

Additional data (n = 226) were obtained from a study conducted during 1995–1996 in Biscayne Bay by the U.S. National Oceanic and Atmospheric Administration [6]. Another data set (n = 218) was obtained from a study of Pearl Harbor

* To whom correspondence may be addressed (ed.long@noaa.gov).

Table 1. Percentage incidence of highly toxic samples and average percentage amphipod survival in marine sediment samples classified according to numerical sediment quality guidelines[a]

Chemical characteristics relative to sediment guidelines	% Highly toxic[b] samples in amphipod survival tests				Average control-adjusted amphipod survival			
	National database[c] (n = 1,068)	Biscayne Bay, FL (n = 226)	Pearl Harbor, HI (n = 219)	Combined summary (n = 1,513)	National database[c] (n = 1,068)	Biscayne Bay, FL (n = 226)	Pearl Harbor, HI (n = 219)	Combined summary (n = 1,513)
Category 1								
Mean ERM quotients <0.1	11	3	4	9	93	95	94	93
Mean PEL quotients <0.1	10	3	6	8	93	95	92	93
No ERLs exceeded	11	5	14	9	92	94	93	92
No TELs exceeded	9	6	0	8	92	93	102	92
Category 2								
Mean ERM quotients 0.11–0.5	30	15	5	21	81	85	97	86
Mean PEL quotients 0.11–1.5	25	23	12	21	84	81	93	86
1–5 ERMs exceeded	32	46	24	32	79	66	86	79
1–5 PELs exceeded	24	37	2	18	83	70	99	88
Category 3								
Mean ERM quotients 0.51–1.5	46	73	61	49	74	48	68	70
Mean PEL quotients 1.51–2.3	50	67	33	49	66	46	76	68
6–10 ERMs exceeded	52	67	71	57	63	47	66	59
6–20 PELs exceeded	47	59	48	48	71	57	73	70
Category 4								
Mean ERM quotients >1.5	75	75	100[d]	76	43	31	44	41
Mean PEL quotients >2.3	77	75	33[e]	73	47	31	56	46
>10 ERMs exceeded	85	75	33[e]	80	41	31	56	41
>20 PELs exceeded	88	75	50[d]	83	38	10	45	37

[a] ERM = effects range—median; PEL = probable effects level; ERL = effects range—low; TEL = threshold effects level.
[b] Mean survival significantly different from controls and <80% of controls.
[c] [3].
[d] 100% (two of two) either marginally toxic ($p < 0.05$, mean survival ≥80% of control) or highly toxic.
[e] 100% (three of three) either marginally toxic ($p < 0.05$, mean survival ≥80% of control) or highly toxic.

conducted in 1997 by the U.S. Navy (unpublished data). These two data sets were selected for evaluation primarily because they were developed with the same methods used in the previous analysis [4] for sample collections, chemical analyses, and amphipod toxicity tests. They were selected also because of the somewhat unusual geochemical properties in both bays. Most of the samples from Biscayne Bay were very sandy and composed primarily of carbonate shell debris. Other samples from adjoining canals were fine-grained silts high in organic carbon content. The Pearl Harbor samples consisted primarily of fine-grained materials accumulated from erosion of iron-rich soils and lava.

Amphipod survival data are reported as percentage of nontoxic controls (average control-adjusted survival). Test results were considered to be nontoxic when mean survival in a sample was not significantly different ($p > 0.05$) from those performed with negative controls. As previously described [3], results were considered as marginally toxic when mean survival was significantly lower than in negative controls ($p < 0.05$) but exceeded 80% of the controls. Highly toxic samples were those in which survival was significantly lower than in controls and <80% of mean survival in controls. The statistical significance of average amphipod survival of less than 80% in classification of sediments as toxic was determined in power analyses [7]. Sediments in which average control-adjusted amphipod survival is below 80% often are defined as toxic in dredging studies [8]. Previously described methods [3] were used to calculate mean SQG quotients.

Statistical methods previously described [3,4] to evaluate the predictive abilities of the SQGs were used in the data analyses. In the previous study [4], we determined the chemical concentrations relative to the SQGs that resulted in four categories of toxicity (i.e., indices of ~10%, ~25%, ~50%, and ~75%). Samples were classified as such with four independent indices: numbers of ERLs/ERMs exceeded, numbers of TELs/PELs exceeded, mean ERM quotients, and mean PEL quotients. Following the previously used procedures [4], we determined the percentages of samples that were highly toxic to amphipods with the new data for each of the four indices of chemical contamination within each of the four classification categories. In addition, the average percentage survival of the amphipods in the toxicity tests was determined for all samples included in each category. Data from the national data set (n = 1,068) were combined with those from Biscayne Bay and Pearl Harbor to provide new estimates (n = 1,513) of the probabilities of observing acute toxicity.

RESULTS

Results of the analyses conducted with the national database indicated that only 9 to 11% of samples were highly toxic in samples classified as Category 1, that is, in which none of the ERLs or TELs were exceeded or mean SQG quotients were <0.1 (Table 1). Average control-adjusted survival in these samples ranged from 92 to 93%, well above the critical threshold of 80% survival.

In the national database, the incidence of highly toxic responses increased sequentially from 9 to 11% in Category 1 to 24 to 32% in Category 2, to 46 to 52% in Category 3, and to 75 to 88% in Category 4 as chemical concentrations increased (Table 1). Average amphipod survival decreased se-

quentially in these samples from 92 to 93%, to 79 to 84%, to 63 to 74%, and to 38 to 47%. It is noteworthy that average survival approximated the critical threshold of 80% in Category 2 samples.

The data from the Biscayne Bay study followed a pattern similar to that of the national database (Table 1). However, the percentages of toxic samples in Category 1 (3–6%) were slightly lower than in the national data base (9–11%). Average survival was less than 80% in samples with 1 to 5 ERMs or PELs exceeded. In addition, as compared to the national database, the percentages of samples in Biscayne Bay that were highly toxic were higher and the average amphipod survival was lower in samples with chemical characteristics equivalent to those in the national database of Category 3.

In the Pearl Harbor data set, both the incidence of toxicity and the average control-adjusted survival were more variable within categories than they were in the national or Biscayne Bay data sets (Table 1). Within Category 1, the incidence of toxicity and average survival varied from 0 to 14% and from 92 to 102%, respectively. Similar variability was apparent in the Category 2, 3, and 4 samples. However, the incidence of toxicity generally increased between categories, and average survival decreased to 66 to 76% in Category 3 and to 44 to 56% in Category 4. Only two or three samples were classified in Category 4, and this reduced our ability to interpret these data. Nevertheless, amphipod survival was at least significantly reduced relative to controls in all the Category 4 samples from Pearl Harbor.

With the data from the three studies combined (n = 1,513), the patterns of both increasing incidence of highly toxic responses and decreasing average survival relative to increasing chemical concentrations were similar to those observed in the national database (Table 1). In samples not expected to be toxic (Category 1) 8 to 9% of samples were highly toxic, and average survival was ≥92%. In the Category 2 samples, however, average incidence of toxicity was somewhat lower and average survival slightly higher than in the national database. Therefore, in the three different data sets, average amphipod survival was either slightly above or below the control-adjusted threshold of 80% in sediments with chemical characteristics equal to those of Category 2. In Category 3 sediments, approx. one-half of the samples were highly toxic, and average survival was ≤70%. In samples with the highest chemical concentrations, 73 to 83% were highly toxic, and average survival was ≤46%.

DISCUSSION AND CONCLUSION

In summarizing his observations of the many attempts by scientists to establish causal links between anthropogenic stresses and effects on biological resources, Sindermann [9] suggested the principal criteria for establishing such relationships were (1) to establish the statistical relationship and demonstrate clear differences between exposed and unexposed populations and (2) to demonstrate that the relationships are consistent over space and time; (3) establish the precision with which the relationship is specific to either species, location, or time; (4) establish that exposure precedes the response, not vice versa; (5) determine that the relationship is plausible and consistent with previous hypotheses; (6) quantify the statistical significance or probability of the relationship; and (7) describe the predictive performance drawn from the observed association. However, he cautioned that satisfaction of all these criteria did not necessarily constitute scientific proof of causality but rather provided a weight of evidence with which to infer a causal relationship.

In this paper, the relationships between the concentrations of chemical substances in sediments and the incidence of toxicity in laboratory tests were quantified using numerical guidelines. The relationships satisfied many of Sindermann's criteria. The data indicated clear differences in toxicity between exposed and unexposed populations, that is, in samples with no SQGs exceeded versus those with many (e.g., >10 ERMs) exceeded. Generally, the incidence of toxicity and average survival changed very little (<5%) by addition of the new data to the national database, demonstrating that the relationships were robust and consistent over space and were not location-specific. Increasing toxicity (response) was related to and tracked with increasing chemical concentrations (exposure), and this relationship constituted a plausible exposure–response association consistent with what had been previously reported. The data assembled in Table 1 summarize the likelihood (probability) of acute toxicity occurring within ranges in chemical concentrations and thus provide a set of predictive tools for assessing the relative quality of sediments.

The data from this study indicate that there is a very low likelihood of acute toxicity in amphipod survival tests and that average survival likely will exceed 80% in sediments with all chemical concentrations below the ERL or TEL values or with mean ERM or mean PEL quotients <0.1. Therefore, the probabilities of misclassifying such samples as acceptable or likely nontoxic or background when they are actually toxic are very small (<10%). In contrast, there is a much higher probability (≥48%) that samples would be toxic in which six or more ERM or PEL values are exceeded or in which mean ERM quotients exceed 0.5 or mean PEL quotients exceed 1.5. Samples with chemical characteristics equivalent to those of Category 2 sediments may or may not be acutely toxic, depending on their specific geochemical characteristics. Therefore, these are sediments with the least certainty of accurately predicting toxicity.

The data indicated that a small degree of variability in toxicological responses that is likely attributable to regional differences in the geochemistry of sediments and the relative bioavailability of sediment-associated toxicants can lead to differences in the predictive abilities of sediment guidelines. Contaminants in the coarse carbonate sands of Biscayne Bay would be expected to be much more bioavailable than those found in the silty clays of other estuaries. The relatively high incidence of toxicity and relatively low percentage survival in Category 2 and 3 samples from Biscayne Bay seemed to substantiate this hypothesis. Users of SQGs should be aware of this regional variability when classifying samples as acceptable and unacceptable with the guidelines. However, despite these slight regional differences in numerical results, the data showed the same pattern in all areas; that is, the probability of acute toxicity generally increased with increasing chemical contamination of the sediments as calibrated to the SQGs. Therefore, we conclude that the sediment guidelines can be used to reliably estimate the probability of acute toxicity in laboratory bioassays.

Acknowledgement—Data from Pearl Harbor were kindly provided by Peter Nakamura (Pacific Division, U.S. Naval Engineering Command, PACNAVFACENGCOM, Pearl Harbor, HI, USA), with assistance from William Lester and John Clayton (Ogden Environmental, San Diego, CA, USA).

REFERENCES

1. Long ER, MacDonald DD, Smith SL, Calder FD. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environ Manag* 19: 81–97.
2. MacDonald DD, Carr RS, Calder FD, Long ER, Ingersoll CG. 1996. Development and evaluation of sediment quality guidelines for Florida coastal waters. *Ecotoxicology* 5:253–278.
3. Long ER, Field LJ, MacDonald DD. 1998. Predicting toxicity in marine sediments with numerical sediment quality guidelines. *Environ Toxicol Chem* 17:714–727.
4. Long ER, MacDonald DD. 1998. Recommended uses of empirically derived, sediment quality guidelines for marine and estuarine ecosystem. *J Hum Ecol Risk Assess* 4:1019–1039.
5. Ingersoll CG, et al. 1997. Workgroup summary report on uncertainty evaluation of measurement endpoints used in sediment ecological risk assessments. In Ingersoll CG, Dillon T, Biddinger GR, eds, *Ecological Risk Assessment of Contaminated Sediments.* SETAC Special Publication. Society of Environmental Toxicology and Chemistry, Pensacola, FL, USA, pp 271–296.
6. Long ER, et al. 1999. Magnitude and extent of chemical contamination and toxicity in sediments of Biscayne Bay and vicinity. NOAA/NOS CCMA 141. Technical Memorandum. National Oceanic and Atmospheric Administration, Silver Spring, MD, USA.
7. Thursby GB, Heltshe J, Scott KJ. 1997. Revised approach to toxicity test acceptability criteria using a statistical performance assessment. *Environ Toxicol Chem* 16:1322–1329.
8. U.S. Environmental Protection Agency, Army Corps of Engineers. 1991. Evaluation of dredged material proposed for ocean disposal. EPA 503/8-91-001. U.S. Testing Manual. Washington, DC.
9. Sindermann CJ. 1997. The search for cause and effect relationships in marine pollution studies. *Mar Pollut Bull* 34:218–221.