J. Great Lakes Res. 22(3):495–511
Internat. Assoc. Great Lakes Res., 1996

A Comparison of Sediment Toxicity Test Methods at Three Great Lake Areas of Concern

G. Allen Burton, Jr.,[1,8] Christopher G. Ingersoll,[2] LouAnn C. Burnett,[3,9] Mary Henry,[4,10] Mark L. Hinman,[5,11] Stephen J. Klaine,[5,12] Peter F. Landrum,[6] Phillipe Ross,[3,13] and Marc Tuchman[7]

[1]*Institute for Environmental Quality*
Wright State University
Dayton, Ohio 45435

[2]*Midwest Science Center*
National Biological Service
Columbia, Missouri 65201

[3]*Illinois Natural History Survey*
Champaign, Illinois 61820

[4]*National Biological Service*
University of Minnessota
St. Paul, Minnessota 55108

[5]*Memphis State University*
Memphis, Tennessee 38152

[6]*Great Lakes Environmental Research Laboratory*
Ann Arbor, Michigan 48105

[7]*Great Lakes National Program Office*
U.S. Environmental Protection Agency
Chicago, Illinois 60604

ABSTRACT. *The significance of sediment contamination is often evaluated using sediment toxicity (bioassay) testing. There are relatively few "standardized" test methods for evaluating sediments. Popular sediment toxicity methods examine the extractable water (elutriate), interstitial water, or whole (bulk) sediment phases using test species spanning the aquatic food chain from bacteria to fish. The current study was designed to evaluate which toxicity tests were most useful in evaluations of sediment contamination at three Great Lake Areas of Concern. Responses of 24 different organisms including fish, mayflies, amphipods, midges, cladocerans, rotifers, macrophytes, algae, and bacteria were compared using whole sediment or elutriate toxicity assays. Sediments from several sites in the Buffalo River, Calumet River (Indiana Harbor), and Saginaw River were tested as part of the U.S. Environmental Protection Agency's (USEPA) Assessment and Remediation of Contaminated Sediments (ARCS) Project. Results indicated several assays to be sensitive to sediment toxicity and able to discriminate between differing levels of toxicity. Many of the assay responses were significantly correlated to other toxicity responses and were similar based on factor analysis. For most applications, a test design consisting of two to three assays should adequately detect sediment toxicity, consisting of various groupings of the following species: Hyalella azteca, Ceriodaphnia dubia, Chironomus riparius, Chironomus tentans, Daphnia magna, Pimephales promelas, Hexagenia bilineata, Diporeia sp., Hydrilla verticillata, or Lemna minor.*

INDEX WORDS: *Great Lakes, fresh water, bioassay, toxicity, sediments.*

INTRODUCTION

Sediment toxicity testing is a recent approach used in ecological risk assessments. The first sediment tests were developed because of concerns over dredged material contamination and its suitability for open-water disposal by the U.S. Army Corps of Engineers (USCOE), in the late 1960s and early 1970s. There was relatively little testing until the 1980's, with a dramatic increase in the past few years (Burton 1991). The science has progressed quickly because of the methodological similarities to the more traditional water column and effluent toxicity tests. The USEPA is developing approaches for managing contaminated sediments and method standardization that will undoubtedly result in an

[8]To Whom Correspondence May Be Addressed
Current Addresses:
[9]University of Illinois Div. of Environ. Health & Safety, Urbana, IL
[10]National Biological Service, Arlington, VA
[11]Exxon Biomedical Sciences, East Millstone, NJ
[12]ENTOX/TIWET, Clemson University, Pendleton, SC
[13]Biology Department, The Citadel, Charleston, SC

even greater amount of sediment testing and research in the near future (Southerland *et al.* 1992).

The objective of a sediment toxicity test is to determine whether sediment is harmful to aquatic organisms. The tests measure interactive toxic effect of complex contaminant mixtures in sediment. These tests do not require knowledge of specific pathways of interactions among sediment and test organisms (Kemp and Swartz 1988). Toxicity testing of sediment can be used to: (1) determine the relationship between toxic effects and bioavailability, (2) investigate interactions among contaminants, (3) determine spatial and temporal distribution of contamination, (4) evaluate hazards of dredged material, (5) rank areas for clean up, and (6) estimate the effectiveness of remediation and management. Sediments spiked with known concentrations of contaminants can be used to establish cause and effect relationships between chemicals and biological responses.

Tests that have been used to evaluate the toxicity of freshwater sediments include: (1) microbial enzyme systems and bacteria, (2) algae, (3) macrophytes, (4) amphipods, (5) midges, (6) mayflies, (7) cladocerans, (8) oligochaetes, and (9) fish (Burton 1991). The choice of the test organism has a major influence on the ecological relevance, success, and interpretation of the test. Furthermore, no one species is best suited for all applications over the wide range of sediment characteristics.

Various methods have been developed to evaluate sediment toxicity. These procedures range in complexity from short-term lethality tests that measure effects of individual contaminants on single species to long-term tests that determine the effects of chemical mixtures on the structure and function of communities. The evaluated sediment phase may include whole sediment, suspended sediment, elutriates, or sediment extracts (Lamberson *et al.* 1992, Burton 1991). For a review of sediment toxicity test methods, their strengths and limitations, and considerations related to sampling and testing of sediments, see Burton (1992).

Currently, there are ASTM standard guides for several of the test species used in this study, e.g., E1706-95c, E1391-94, E1525-94a. In addition, the USEPA has recently standardized test methods for *Hyalella azteca*, *Chironomus tentans,* and *C. riparius*, and the bioaccumulation assay using *Lumbriculus variegatus* (USEPA 1994). The U.S. Army Corps of Engineers and the USEPA also have test methods for dredged material evaluations (USEPA-USCOE 1994). These standard test procedures may vary slightly from those used in this ARCS study.

A primary objective of the ARCS study was to determine which toxicity assays efficiently identified sediment contamination in the Great Lakes. In the ARCS study, a multi-biological level approach was taken, using assays representative of multiple trophic levels which had been successfully used in studies of sediment contamination.

METHODS

A detailed description of the toxicity test methods is presented in Burton *et al.* (1989) and Ingersoll *et al,* (1993). Sample site descriptions and sediment collection methods are described in detail in Ingersoll *et al.* (1993). Briefly, sediments were collected using a ponar sampler. A minimum of 5 composites were used for each station and mixed using a cement-mixer. Sediments were subsampled following mixing and distributed to high density polyethylene bottles (2 to 4 L). Sediments were kept in ice chests on ice until sampling was completed (within 2 days) and then shipped via overnight express to the participating laboratories. Upon sample receipt, sediments were refrigerated until test initiation (typically within 2 weeks of receipt). A Quality Assurance Project Plan was developed, which included reference toxicant testing, laboratory audits and data checks.

The sampled sediments and elutriates were tested with a variety of species. Toxicity assays were conducted with organisms from in-house cultures (exception was field-collected *Hexagenia* and *Diporeia*), and purchased rotifer cysts and Microtox. Culture health was routinely monitored by measuring reproductive rates and conducting reference toxicant testing. Test species included: (1) fathead minnow (*Pimephales promelas*, < 24 h old, whole sediment), (2) cladoceran (*Daphnia magna* and *Ceriodaphnia dubia,* < 24 h old, elutriates or whole sediment) neonates, (3) amphipods (*Hyalella azteca*, 7–14 d old, and *Diporeia* spp., adults, (formerly *Pontoporeia hoyi*), whole sediment), (4) midges (*Chironomus riparius* and *Chironomus tentans*, 8–12 d, second to third instar, whole sediment), (5) mayflies (*Hexagenia bilineata*, juvenile instars, elutriates, and whole sediment), (6) duckweed (*Lemna minor,* mature plants, whole sediment), (7) macrophyte (*Hydrilla verticillata*, mature plants, whole sediment), (8) rotifers (*Brachionus calciflorus*, < 24 h old, elutriates), (9) microbial enzymes (whole sediment, elutriates) and Microtox (elutriates), and (10) algae (*Selenastrum capricornutum*, log-phase growth, elutriates) (Table 1). In situ colonization of artificial substrates (3M Corp.

TABLE 1. Sediment toxicity tests evaluated in the ARCS program.

Biological Level	Test Organism/Community	Duration	Endpoints	Phase
Fish	*Pimephales promelas*	7 day	Larval survival/weight	S
		7 day	Embryo-larval survival, length, terata	S
Zooplankton	*Daphnia magna*	48 hour	Survival	S,E
		7 day	Survival/Reproduction (3 brood)	S
	Ceriodaphnia dubia	7 day	Survival/Reproduction (3 brood)	E
	Brachionus sp	24 hour	Survival	E
Benthic Invertebrate	*Hyalella azteca*	7 day	Survival	E
		14, 28 day	Survival, length, antenna segment number, sexual maturation	S
	Diporeia spp.	28 day	Survival	S
		5 day	Preference/avoidance, survival	S
	Chironomus tentans	10 day	Survival, length/weight	S
	Chironomus riparius	14 day	Survival, length/weight	S
	Hexagenia bilineata	10 day	Survival	S,E
			Molting frequency	S,E
	Rapid Bioassessment III (artificial substrates)	(10)	Community indices	S
Phytoplankton	*Selenastrum capricornutum*	48, 96 hour	Growth	E
		24 hour	14°C uptake	E
Macrophyte	*Lemna minor*	4 day	Growth (frond number)	S
			Chlorophyll a	S
			Biomass (wet weight)	S
	Hydrilla verticillata	10 day	Chlorophyll a	S
			Dehydrogenase activity	S
			Shoot length	S
			Root length	S
			Peroxidase	S
Microbial	Microtox ® (*Photobacterium phosphoreum*)		Luminescence	E
	Alkaline phosphatase (sediment community)		Enzyme activity	S
	Dehydrogenase (sediment community)		Enzyme activity	S
	β-Galactosidase (sediment community)		Enzyme activity	S
	Glucosidase (sediment community)		Enzyme activity	S

Note: E – elutriate
S – whole sediment

Summary: Total toxicity test types – 25
Single species tests – 20
Community tests – 5
Total endpoints – 55 (duplicate endpoints in solid and elutriate phases, counted as one)
Single-species endpoints – 41

polymer coiled web mesh) by benthic invertebrates at each site was also evaluated.

With the exception of Microtox, all assays were conducted in replicate. Most assays were conducted with ten to twenty organisms in triplicate test beakers, however, the mayfly, *C. tentans*, and short-term chronic cladoceran assays were conducted with one organism per 10 replicates.

Because of the large water volumes required for this test battery, the decision was made by the GLNPO ARCS Toxicity-Chemistry Work Group to test elutriates instead of interstitial water. All laboratories processed their elutriates in the same manner. A 1:4 mixture of sediment to test water was shaken for 30 minutes and then allowed to settle for 1 hour. The supernatant was then centrifuged for 15 min at 10,000 × g to reduced suspended solids. The centrifuged sample was used for testing. However, in the algal growth assay, elutriates were also filtered via a 0.45 micron cellulose-acetate filter to remove any remaining particulates.

Whole-sediment toxicity assays were conducted with macrobenthos in static or water-renewal systems at temperatures of 20 to 25°C. Sediments were placed in the test beakers and overlying water was gently added, typically resulting in a sediment to overlying water ratio of 1:4. Test organisms were randomly added within 24 h and the test initiated. Exposure water was typically moderately hard (hardness 134 mg/L as $CaCO_3$; alkalinity 1.2 to 1.3 mM 60 to 65 mg/L as $CaCO_3$); pH 7.8 to 8.0; conductivity 300 μmhos/cm; sulfate 72 mg/L). Dissolved oxygen, temperature, alkalinity, pH, conductivity, and water hardness were measured either daily or at test initiation and termination, depending on the parameter. See Burton *et al.* (1989) and Ingersoll *et al.* (1993) for additional detail.

Data Analysis

Data are summarized several different ways to allow for evaluations of the following questions: (1) does Station (sample) A differ from Station B?; (2) does Site 1 differ from Site 2?; (3) how sensitive are the measured endpoint responses?; and, (4) how well does the assay discriminate between different levels of toxicity (does Treatment A differ from Treatment B)? Data were analyzed using replicate data, arithmetic means of each assay by station, arithmetic means of each assay by site, or arithmetic means over all sites (using both station means and replicate data).

By conducting all assays on split samples which were collected and processed in the same manner, statistical relationships could be established with high statistical confidence. The assay responses were evaluated and compared by several methods, as described below:

- Sensitivity: comparison of the test response to the control response. Responses were then grouped into categories, (1) 20 to 50% difference and (2) 20 to 100% difference from the control. A relatively insensitive assay would show small differences from the control (e.g., near 20%) and sensitive assays would show large differences (e.g., greater than 50%). The number of responses falling within each grouping was then used to compare the relative sensitivity between assays. In addition, the variability and range of these responses were determined.
- Discrimination: the ability of the assay to detect differing degrees of toxicity between samples. It is important when defining the spatial extent of contamination to be able to ascertain whether sediment samples vary in toxicity. A statistical test was conducted (Kruskal-Wallis) which was used to determine the degree to which the sediment samples (each station) were different from each other within a test site (e.g., Buffalo River). The more statistically significant the difference was between samples, the more discriminatory the assay was.
- Redundancy: The degree of similarity of the responses from different assays was measured using correlation analyses (parametric and non-parametric) and by grouping the assay responses into patterns, through factor analysis. A high degree of correlation or pattern (grouping) similarity implies that the assays were responding in a similar manner. These analyses were conducted across all sites to better meet the study objective of determining which assays were best (in terms of predictive power) for Great Lakes studies. If the assays are producing similar information then it is less important that each be conducted, unless a weight-of-evidence approach is being used. It is, perhaps, of greater importance that a range of assays be employed which respond differently to varying types of sediment contamination (i.e., which show different response patterns and groupings). This approach will increase the likelihood that any detrimental effects on the aquatic ecosystem are being detected.

Data analyses were made using appropriate parametric or non-parametric correlation and mean

comparison analyses. Correlation analyses, sensitivity determinations and groupings, discriminatory analyses, and principal component analysis were generated using a SAS computer package.

Raw data and summary statistics are presented in Burton *et al.* (1989) and Ingersoll *et al.* (1993). The data have also been entered into the USEPA's Office of Marine and Estuarine Protection ODES database and have received a quality assurance validation from the USEPA (USEPA 1994a).

RESULTS AND DISCUSSION

The principal objectives of the study were to determine: (1) which assays were the most sensitive to the presence of contaminated sediments, (2) which assays discriminated best between differing degrees of sediment contamination, and (3) were the assays responding in similar (e.g., significantly correlated) or dissimilar patterns. As each test sampling was a unique sampling event, sediment toxicity varied widely between stations and sites. Therefore, the degree of assay sensitivity and comparability with other assays varied with each sampling event. The data summaries provide a weight-of-evidence approach for identifying trends in assay responses. It is likely that conclusions from this study may change with additional testing of contaminated sites.

Combined Site Sediment Toxicity—Sensitivity

The assays were ranked by comparing each assay replicate to the control and grouping response into: (1) effects greater than 50% difference from the control and (2) effects between 20 and 50% different from the control (Table 2). Groupings were divided into inhibitory and stimulatory responses. In the "Effect Level" column of Table 2, the percentage of samples classified as toxic are listed. They are then ranked by their relative degree of toxicity (% effect level) in the "Site Rank" column. The assay responses were ranked against each other. Equal effect levels received equal rankings. The "Composite Rank" was the average mean value of the four site rank values (with the associated range).

Several benthic species were very responsive to sediment toxicity. Preference behavior by *Diporeia* was the most sensitive endpoint of effects, being affected by 90% of the samples. Behavior would be expected to be a responsive sublethal measure, but the ecological significance of a behavioral response is difficult to interpret. *Hexagenia bilineata* endpoints comprised 4 of the top 15 inhibitory measures, with survival and molting being affected in both elutriate and whole-sediment exposures. Unfortunately, the number of data points used in the mayfly assays was only 16 and many of these samples were stored for several months before testing which makes comparisons to *H. bilineata* suspect. The *Hydrilla verticillata* assay produced primarily stimulatory responses with a maximum of 88% of the samples being affected in the peroxidase assay and 75% of the samples inhibiting root growth (and stimulated in 10% of the samples).

The 11 assays ranked in Table 2 comprised 43 response endpoints. The remainder of the assays and endpoints were deleted from this ranking because there was either insufficient data or the controls were not appropriate for the sensitivity calculation used in the ranking process (e.g, microbial enzymes or artificial substrate colonization). Only the *H. verticillata* peroxidase and the *H. azteca* 14 d antenna length endpoints did not show any test responses in greater than 50% of the samples.

Combined Sediment Toxicity—Discrimination

A ranking of the discriminatory power of 53 of the 97 endpoints is listed in Table 3. The discriminatory ability measures how well the assay response detects varying levels of sediment toxicity. This ability is expressed using levels of statistical significance, or p values. The smaller the p value, the higher the level of statistical significance, or difference between samples/stations. Some assay data were not available or could not be analyzed by this procedure, so p values do not exist for all responses at all sites. The p value average, the range of p values, and the number of sites (1 to 4) for which discrimination analysis was conducted, must all be considered in the relative ranking. It is misleading, in some cases, to only consider the p value average, if it only came from one site or highly significant p values (e.g., $p = 0.0001$) were offset by very high p values (e.g., $p = 1.0$). For example, the first two endpoints were not considered reliable as they were only reported for one site.

The photosynthetic and indigenous microbial responses would be expected to be good discriminators since they can exhibit both inhibitory and stimulatory responses, giving them a wider range of response than just 0 to 100%, as with conventional test responses. Indeed, the *S. capricornutum* growth at 48 h (average p value of 0.0213) and at 96 h (average p value of 0.0150) were among the best discriminatory

TABLE 2. Ranking of most sensitive assay endpoints.[a]

Assay[b]	Rank	Effect Level[c] 20–100%	Effect Level[c] 20–50%	Site[d] Ranks IH	Site[d] Ranks BR	Site[d] Ranks SR1	Site[d] Ranks SR3	Composite Avg. (Range)
php.s	1	90	6	1=[e]	3	5	1	2.5 (1–5)
xs.e	2	75	0	1=	1	1	20=	5.75 (1–20)
vr.s	3	75 (10)[f]	35	1=	11=	3	2	4.25 (1–11)
xm.e	4	69	13	1=	4	2	20=	6.75 (1–20)
dr.s	5	64	44	6	7	6	8	6.75 (6–8)
Hx14	6	60 (23.3)	1.7	1=	2	17	7	6.75 (1–17)
flw.s	7	58	34	3	8	8	10	7.75 (3–10)
M45.5	8	54	12	2	9	23=	–	11.3 (2–23)
hs.s	9	51	16	1=	5	10	17=	8.25 (1–17)
M45.15	10	50	4	1=	10	23=	–	11.3 (1–23)
Ct10	11	50	12	1=	26=	–	4	10.3 (1–26)
cr.s	12	50	13	10=	18	4	3	8.75 (3–18)
xs.s	13	50	19	1=	17	19=	6	10.75 (1–19)
Cr14	14	47	10	1=	6	15	19=	10.25 (1–19)
xm.s	15	44	6	1=	11=	19=	20=	12.75 (1–20)
phs.s	16	43	27	14	14	7	9	11.6 (7–14)
Hs14	17	35.5	9.2	1=	13	19	20=	13.25 (1–20)
Ctl	18	30.8	30.8	5	12	–	20=	12.3 (5–20)
fet.s	19	31	21	4	17	11	20=	13.0 (4–20)
cr.e.100	20	31	3	7	26=	14	14	15.25 (7–26)
vs.s	21	29 (52)	29	9	23	13	15	12.5 (5–23)
cs.s	22	26	0	16	21	20	11	17.0 (11–21)
fes.s	23	21	13	8	22	27=	17=	18.5 (8–27)
ds.48.s	24	19	9	12	26=	12	20=	17.5 (12–23)
cs.e.100	25	18	0	17	24	16	16	18.25 (16–24)
vd.s	26	17.5 (70)	5	21=	16	25	5	16.75 (5–25)
Hs28	27	15	5	–	–	19=	18	18.5 (18–19)
vc.s	28	15 (53)	11	20=	20	9	20=	17.75 (9–20)
M100.5	29	14	7.4	–	26=	28=	12	22.0 (12–28)
fls.s	30	14.5	9.2	13	19	28=	19=	19.75 (13–28)
Hl14	31	13.3	13.3	10=	15	25	20=	17.5 (10–25)
lb.s	32	12 (36)	1	11	21	28=	20=	20.0 (11–28)
M100.15	33	11.1	3.7	–	26=	27=	13	22.0 (13–27)
Hx28	34	10 (2.5)	5	–	–	20	19=	19.5 (19–20)
Hl28	35	10	7.5	–	–	18	20=	19.0 (18–20)
ds.s	36	8	0	18	25	24	19=	21.5 (18–25)
fel.s	37	7	7	15	26=	28=	20=	22.25 (15–28)
Crl	38	6.6 (16.7)	3.3	–	26=	17	20=	21.0 (17–26)
Ha28	39	5	5	–	–	21	20=	20.5 (20–21)
lf.s	40	3 (36)	3	19=	26=	28=	20=	23.25 (19–28)
lc.s	41	2 (25)	2	20=	26=	28=	20=	23.25 (20–28)
Ha14	42	0	0	21=	26=	27=	20=	23.5 (20–27)
vp.s	43	0 (87.5)	0	21=	26=	28=	20=	23.75 (20–28)

[a] Ranking based on SAS sensitivity groupings using individual replicate data (not station means). Missing endpoints were not included due to lack of true control values for determining sensitivity (Response – Control/Control) or data were too limited.

[b] See facing page.

[c] Percentage of values showing effects ranging from 20 to 100 % or 20 to 50 % of the control response.

[d] IH, Indiana Harbor; BR, Buffalo River; SR1; Saginaw River survey no. 1; SR#, Saginaw River survey no. 3. Based on mean value of station replicates.

[e] = , Same rank level as another measurement endpoint.

[f] (x) = additional percentage of responses stimulated greater than 20% over the control response. Value not considered in ranking.

Footnote b, Table 2, continued from facing page.

Test Response Endpoints and Codes *.

Code	Response
a.as	amphipods, % (artificial substrate)
ap	alkaline phosphatase activity
BR1	Buffalo River, first sample period
c	control
c14.e.50	*Selenastrum capricornutum* C-14 uptake (24 h 50% elutriate)
c.as	chironomids, % (artificial substrate)
cr	*Ceriodaphnia dubia* reproduction (7 d, 3 brood)
cs	*Ceriodaphnia dubia* survival (7 d, 3 brood)
cr.e.100	*Ceriodaphnia dubia* reproduction (7 d 100% elutriate)
cs.e.100	*Ceriodaphnia dubia* survival (7 d 100% elutriate)
Cr14	*Chironomus riparius* survival (14 d)
Crl	*Chironomus riparius* length (14 d)
Ct10	*Chironomus tentans* survival (10 d)
Ctl	*Chironomus tentans* length (10 d)
dh.s	dehydrogenase activity in sediment
dr	*Daphnia magna* reproduction (7 d, 3 brood)
ds	*Daphnia magna* survival (7 d, 3 brood)
ds48	*Daphnia magna* survival (48 h exposure)
.e	elutriate phase exposure
f.as	flatworms, % (artificial substrate)
fel	*Pimephales promelas* embryo larval length (7 d)
fes	*Pimephales promelas* embryo larval survival (7 d)
fet	*Pimephales promelas* embryo larval terata (7 d)
fls	*Pimephales promelas* larval survival (7 d)
flw	*Pimephales promelas* larval weight (7 d)
ga	galactosidase activity
gl	glucosidase activity
h.as	hydra numbers, % (artificial substrate)
hs7	*Hyalella azteca* survival (7 d)
Hs14	*Hyalella azteca* survival (14 d)
Hl14	*Hyalella azteca* length (14 d)
Ha14	*Hyalella azteca* antenna length (14 d)
Hx14	*Hyalella azteca* sexual maturation (14 d)
Hs28	*Hyalella azteca* survival (28 d)
Hl28	*Hyalella azteca* length (28 d)
Ha28	*Hyalella azteca* antenna length (28 d)
Hx28	*Hyalella azteca* sexual maturation (28 d)
ic.m	Invertebrate Community Index mayfly, % (artificial substrate)
ic.to	Invertebrate Community Index tolerant organisms, % (artif. substr.)
ic.tr	Invertebrate Community Index, taxa richness (artif. substr.)
IH1	Indiana Harbor, first sample period
lb	*Lemna minor* biomass (4 d)
lc	*Lemna minor* chlorophyll a (4 d)
lf	*Lemna minor* frond number (4 d)
m.e.50	Microtox 50% dilution (15 min)
mr.e.50	Microtox QA check on stored rotifer samples, 50% dilution (elutriate)
M100.5	Microtox 100% (5 min)
M100.15	Microtox 100% (15 min)
M45.5	Microtox 45% dilution (5 min)
M45.15	Microtox 45% dilution (15 min)
o.as	oligochaete number (artificial substrate)
php	*Pontoporeia hoyi (Diporeia spp.)* preference (5 d)
phs	*Pontoporeia hoyi (Diporeia spp.)* survival (20 d)
r2d	Rapid Bioass. Protocol Phase II, % contrib. dominant family (artif. subs.)
r2.ec	Rapid Bioassessment Protocol Phase II, EPT/Chironomidae (artif. subs.)
r2f	Rapid Bioassessment Protocol Phase II, Family Biotic Index (artif. subs.)
r2.tr	Rapid Bioassessment Protocol Phase II, taxa richness (artif. substr.)
r.e.50	rotifer (*Brachionus sp.)* survival 24 h 50% elutriate)
.s	solid phase exposure
s.48.e.100	*Selenastrum capricornutum* growth (48 h 100% elutriate exposure)
s.96.e.100	*Selenastrum capricornutum* growth (96 h 100% elutriate exposure)
SR1	Saginaw River, first sample period
SR3	Saginaw River, third sample period
sag.b	Saginaw R. macroinvertebrate biomass (artificial substrate)
vc	*Hydrilla verticillata* chlorophyll
vd	*Hydrilla verticillata* dehydrogenase
vp	*Hydrilla verticillata* peroxidase
vr	*Hydrilla verticillata* root length
vs	*Hydrilla verticillata* shoot length
xm	*Hexegenia bilineata* molting frequency (10 d)
xs	*Hexegenia bilineata* survival (10 d)
z.as	zebra mussel number (artificial substrate)
*	Some codes in the dataset are combinations of the above codes, such as:

c (control) + assay code + .s (solid phase) or + .e (elutriate) + # (% sample, e.g., 100, 71.4, 57, 50, 28.6, 15, 12.5, 10.2 or 6.25). Assays which are capitalized were conducted by the USFWS (National Biological Service).

TABLE 3. Discriminiatory table.

Toxicity Test (Endpoint)[a]	Discriminatory Rank[b]	Average *P* Value[c]	Standard Deviation	Significant Surveys[d]	Range of *P* Values
Hydra (numerical percent, artificial substrate)	1	0.0069	N.A.[e]	1/1	
Saginaw River macroinvertebrates (biomass, artificla substrate)	2	0.0069	N.A.[e]	1/1	
Brachionus sp. (50-percent elutriate, 24-hour survival)	3	0.0071	0.0048	4/4	0.0018-0.0134
Ceriodaphnia dubia (100-percent elutriate, 7-day reproduction)	4	0.0083	0.0110	4/4	0.0001-0.0233
Chironomus riparius (14-day length)	5	0.0116	0.0050	3/3	0.0063-0.0162
Selenastrum capricornutum (100-percent elutriate, 96-hour growth)	6	0.0150	0.0097	4/4	0.0037-0.0273
Sediment microbial community (dehydrogenase activity)	7	0.0152	0.0170	2/2	0.0032-0.0273
Pimephales promelas (7-day larval weight)	8	0.0198	0.0180	4/4	0.0061-0.0463
Selenastrum capricornutum (100-percent elutriate, 48-hour growth)	9	0.0213	0.0155	3/3	0.0084-0.0329
Rapid Bioassessment Protocol Phase II (Family Biotic Index, artificial substrate)	10	0.0240	0.0113	3/3	0.0291-0.0319
Hyalella azteca (28-day length)	11	0.0298	0.0176	3/3	0.0129-0.0481
Sediment microbial community (glucosidase activity)	12	0.0331	0.0314	2/3	0.0050-0.0670
Selenastrum capricornutum (50-percent elutriate, 24-hour 14°C uptake)	13	0.0507	0.0736	3/4	0.0013-0.0158
Daphnia magna (7-day reproduction)	14	0.0570	0.1136	3/4	0.0001-0.2274
Sediment microbial community (galactosidase activity)	15	0.0647	0.0529	2/2	0.0273-0.1021
Flatworms (numerical percent, artificial substrate)	16	0.0675	0.0726	2/3	0.0223-0.1513
Lemna minor (4-day cholrophyll a)	17	0.0676	0.0547	3/4	0.0086-0.1266
Amphipods (numerical percent, artificial substrate)	18	0.0707	0.0488	1/3	0.0284-0.1223
Pimephales promeles (7-day embryo larval terata)	19	0.0826	0.1404	3/4	0.0020-0.2929
Rapid Bioassessement Protocol Phase II (percent contributing dominant family, artificial substrate)	20	0.0870	0.0965	2/3	0.0302-0.1984
Microtox (50-percent dilution, 15 minute)	21	0.0890	0.0060	0/3	0.0833-0.1017
Hyalella azteca (14-day survival)	22	0.1049	0.1441	2/3	0.0173-0.2712
Hydrilla verticillata (10-day peroxidase)	23	0.1051	0.0000	0/1	
Hyalella azteca (28-day survival)	24	0.1098	0.1137	1/3	0.0169-0.2366
Hyrdrilla verticillata (10-day shoot length)	25	0.1182	0.0270	0/4	0.0922-0.1479
Oligochaetes (number, artificial substrate)	26	0.1397	0.2051	2/3	0.0116-0.3763
Rapid Bioassessement Protocol Phase II (taxa richness, artificial substrate)	27	0.1407	0.0977	1/3	0.0290-0.2107

Ceriodaphnia dubia (7-day survival)	28	0.1452	0.1698	2/4	0.0001-0.3402
Hyalella azteca (28-day sexual maturation)	29	0.1639	0.2102	1/3	0.0463-0.5296
Sediment microbial community (alkaline phosphatase activity)	30	0.1712	0.2671	2/3	0.0032-0.4712
Hyalella azteca (28-day antenna segment number)	31	0.1726	0.2391	2/3	0.0219-0.4483
Hyalella azteca (14-day length)	32	0.1805	0.1372	1/3	0.0277-0.2930
Pimephales promelas (7-day embryo larval length)	33	0.1808	0.1928	2/4	0.0183-0.3766
Ceriodaphnia dubia (7-day reproduction)	34	0.1914	0.3613	3/4	0.0002-0.7329
Ceriodaphnia dubia (elutriate, 7-day survival)	35	0.1930	0.2234	2/4	0.0001-0.4060
Lemna minor (4-day biomass)	36	0.2017	0.2509	1/4	0.0452-0.5743
Chironomus riparius (14-day survival)	37	0.2091	0.3290	3/4	0.0245-0.7017
Daphnia magna (7-day survival)	38	0.2441	0.2865	2/4	0.0001-0.5527
Diporeia spp. (5-day preference)	39	0.2671	0.2441	2/4	0.0539-0.8296
Daphnia magna (48-hour survival)	40	0.2764	0.2212	1/4	0.0272-0.4060
Hyallela azteca (14-day sexual maturation)	41	0.2765	0.2425	1/3	0.0463-0.5296
Hyalella azteca (14-day antenna segment number)	42	0.3120	0.3849	1/3	0.0262-0.7496
Pimephales promelas (7-day larval survival)	43	0.3172	0.4498	2/4	0.0161-0.9716
Diporeia spp. (28-day survival)	44	0.3182	0.4398	2/4	0.0233-0.9576
Hydrilla verticillata (10-day chlorophyll)	45	0.3720	0.1783	0/4	0.1931-0.6110
Chironomids (numerical percent, artificial substrate)	46	0.3633	0.4987	1/2	0.0105-0.7161
Pimephales promelas (7-day larval survival)	47	0.3815	0.3122	1/4	0.0159-0.7580
Lemna minor (4-day frond number)	48	0.3824	0.3065	1/4	0.0427-0.7863
Rapid Bioassessment Protocol Phase II (EPT/Chironomidae, artificial substrate)	49	0.4191	0.5078	0/3	0.0593-1.0000
Hyalella azteca (7-day survival)	50	0.4742	0.00000	0/4	0.0713-1.0000
Hydrilla verticillata (10-day dehydrogenase)	51	0.4988	0.3647	0/3	0.1125-0.8371
Zebra mussels (numbers, artificial substrate)	52	0.5080	0.6958	1/2	0.0160-1.0000
Hydrilla verticillata (10-day root length)	53	0.5826	0.3152	0/4	0.2521-0.8769

[a]All toxicity tests were conducted with whole sediment unless indicated otherwise. Some endpoints lacked adequate data for ranking.
[b]Discriminatory ranks based on the average P value for pairwise statistical comparisons of all station responses with the control response.
[c]Average of P values for the surveys analyzed.
[d]Number of surveys with significant Kruskal-Wallis P value ($P \leq 0.05$) per total number of surveys where that endpoint was analyzed.
[e]Not applicable since standard deviation cannot be done on one sample.

assays over four sites. However, of the other photosynthetic endpoints, only *L. minor* chlorophyll a production showed significant differences at three sites. Frond number and biomass showed differences at only one site and *H. verticillata* endpoints did not detect any significant differences.

The indigenous microbial responses were better discriminators than these later two photosynthetic surrogate species with significant differences observed at two or three of the three sites analyzed. They ranked from high to low discriminatory ability, in order, as: dehydrogenase, glucosidase, galactosidase, and alkaline phosphatase.

Several of the benthic macroinvertebrate community indices that were sampled using artificial substrates were good discriminators. The top two listed in Table 2 (hydra and biomass) cannot be reliably evaluated as they only were analyzed or determined for one site. The Family Biotic Index, however, was highly discriminatory ($p = 0.0291$ to 0.0319) at all three sites. The second best discriminator in this group of endpoints was percent flatworm composition, showing differences at two of three sites. Two other endpoints showing this level of discrimination, but with slightly lower p values, were percent contributing dominant family and percent oligochaete composition.

Of the other toxicity test species which were evaluated, several endpoints were good discriminators among the nonbenthic species. Survival of the benthic species *Hyalella azteca, C. riparius*, and *Diporeia* did not rank high in discriminatory ability at the four sites. However, chronic endpoints of length and sexual maturation were highly discriminatory at a minimum of one test site. The *C. riparius* length ($p = 0.0116$) and *H. azteca* 28 d length ($p = 0.0298$) were significant at all three sites tested. The best nonbenthic invertebrate endpoints were ranked as follows: the rotifer, *Brachionus* survival, *C. dubia* reproduction (elutriate), and *P. promelas* larval weight, showing differences at all four sites. The rotifer assay showed significant discrimination at all four sites, however, the data are questionable, for comparison purposes, due to storage of sediment for 12 months before testing. Five endpoints had significant p values at 75% of the sites, including: *S. capricornutum* ^{14}C-uptake, *D. magna* reproduction, *P. promelas* embryo- larval terata (visible abnormalities), *C. dubia* reproduction (whole sediment), and *C. riparius* survival. Some other endpoints showed highly significant p values at two of four sites, but had high p values at others, such as *C. dubia* survival (whole sediment), *P. promelas* embryo larval length, *C. dubia* reproduction (whole sediment) and survival (elutriate), and *D. magna* survival (whole sediment). In summary, there were several assay endpoints which proved to be highly discriminatory of degrees of sediment toxicity. This is a critically important assay trait when attempting to define the spatial extent of site contamination. The nonbenthic assays tended to be more discriminatory than the benthic assays and should be included in any test battery for this reason.

Summary of Assay Sensitivity and Discriminatory Ability

While the *Diporeia* preference and avoidance endpoints were the most sensitive overall, this assay is one of the least developed (Gossiaux *et al.* 1993). The survival endpoint for this organism was relatively insensitive (ranked from 7 to 14 in the four surveys) and *Diporeia* must be collected from the field for testing. The ecological significance of behavior endpoints, such as avoidance/preference is difficult to evaluate at this time. However, *Diporeia* is of critical importance in the Great Lakes. This characteristic alone indicates the assay should be given high priority for additional methods development and testing.

Hexagenia bilineata exhibited sensitive responses at most test sites with endpoint responses ranging from 1 to 9. The Kruskal Wallis test could not be run with this data set. Previous discriminatory analysis using a different procedure whereby the geometric mean is divided by the arithmetic mean indicated the molting endpoint to be relatively discriminatory (rank = 5), however survival was not discriminatory (rank = 21). Surprisingly, the elutriate exposures were, for *H. bilineata*, more sensitive than the whole-sediment assays. The sensitivity of *H. bilineata* in the ARCS project may have resulted from the prolonged storage of sediment before testing. The inability to continuously culture mayflies in the laboratory has limited their routine use in sediment testing. Mayflies may also be sensitive to grain size in whole-sediment exposures (ASTM 1995).

An elutriate assay was to be evaluated with the nematode *Pangrellus redivivus* (Samoiloff *et al.* 1980). The culture was lost during the project. Consequently the rotifer, *Brachionus calyciflorus,* survival assay was testing in place of nematodes (Snell and Persoone 1989). Unfortunately, the assays had to be conducted after prolonged sediment storage (up to 12 months). As with the *Hexagenia* assay, comparison of sediment effects on rotifers to the other assays is

tenuous because of potential toxicity artifacts caused by sediment storage. The rotifer was insensitive but was discriminatory in elutriate exposures.

Hyalella azteca responses were highly variable, depending on the time of exposure (7 to 28 d) and the endpoint measured, with rank levels ranging from 1 to 27. The advantages to conducting sediment tests with *H. azteca* are: (1) the animals can be cultured in the laboratory, (2) testing and culturing methods have been standardized, (3) effects on survival, growth, or sexual maturation can be monitored in 7- to 28-d exposures, (4) *H. azteca* are insensitive to grain size of the sediment (Ankley *et al.* 1996), (5) *H. azteca* ranked number 4 in the combined ranking of sensitivity and discriminatory ability for 14 d survival, and (6) they correlated well with other assay responses and covered three of the four unique pattern groupings identified over all the test sites.

The midges *C. tentans* and *C. riparius*, as *H. azteca*, showed a wide range of sensitivity and discriminatory ability over the test sites, but ranked relatively high overall. Control survival for the *Chironomus* sp. was typically lower than the other test species. The advantages to conducting sediment tests with midges are: (1) the animals can be cultured in the laboratory, (2) testing and culturing methods have been standardized, and (3) effects on survival and growth can be monitored in 10- to 14-d exposures. The *C. riparius* assay responses described two of the four unique pattern groupings observed at the ARCS sites.

Tests with the aquatic macrophyte *Hydrilla* have been conducted by very few laboratories. Some of the measured endpoints used in this assay proved to be sensitive (root growth, ranked 1–11), but the endpoints were not discriminatory. *Hydrilla* represents a unique level of biological organization and should be considered in future assessments if adequate resources are available for testing. The *Lemna* (duckweed) assay also measures a unique biological level of organization that is of importance to ecosystem functioning. The assay by design cannot be highly sensitive to sediment contaminants since the plants float on the surface of the water. Therefore, *Lemna* are only exposed to contaminants that are water soluble or associated with colloidal suspended particles.

Hall *et al.* (1996, this issue) reported problems conducting elutriate toxicity tests using the 24-h ^{14}carbon assimilation with algal assay. Interpretations of toxicity using *Selenastrum capricornutum* were complicated by variable nutrient and inorganic carbon concentrations in the elutriate samples. All of the elutriate samples tested stimulated carbon assimilation of *Selenastrum capricornutum* in one or more of the dilutions. Attempts to modify the algal medium to provide unlimited nutrients were not successful. An algal medium that supports greater growth potential should be developed in order to evaluate the toxicity of environmental samples with high concentrations of nutrients to algae.

The Microtox assay response was relatively sensitive (overall rank of 8; Table 2). It was well correlated with other assay responses. Other strengths of the Microtox assay are rapid response and small volume requirements. Microtox is recommended as a screening tool for evaluating freshwater and marine sediment, and can be used to quickly process a large number of samples. The bioluminescence response is relatively sensitive to many contaminants and is often correlated with responses of other test species. The simplicity, degree of standardization, and ability to test interstitial water, elutriates, and whole sediment makes Microtox a versatile and relevant assay. However, no one assay should be used as the sole toxicity method, since the assay can produce false negatives or false positives.

The indigenous assays included benthic macroinvertebrate indices from artificial substrates and microbial enzyme activities of sediment samples. These data could not be analyzed for sensitivity with the above data sets because of the lack of controls for comparisons. Several endpoints for these assays proved to be highly discriminatory (Table 3). The percent tolerant species and percent chironomid composition indices were highly correlated with toxicity assay responses. Both of these assays represent unique levels of biological organization. Microbial enzyme and benthic colonization assays evaluate indigenous organisms, not surrogate species and, therefore, there are reduced uncertainty in data extrapolations.

Measured Endpoint Response Pattern Similarities–Redundancy

Principal component analysis (PCA) was conducted on the data set to determine if there were similar groupings of responses among assays. This procedure illustrates which endpoints produce similar or unique response patterns. In the PCA, the data undergo an orthogonal transformation, so the factors are independent of each other. The results of the analysis are presented as separate factors, each of which explains one response pattern. The percent contribution of each variable (endpoint) to each re-

sponse pattern (factor) is listed in Table 4. Variables contributing similar levels to a factor are grouped as being similar. There can be no missing data for any variable; that is, the number of data points must be equal. So, there were only 20 endpoints (Table 4) out of 97 possible endpoints that met these data requirements.

The statistical analysis revealed four groupings of unique response patterns. Group 1 consisted of *H. azteca* length (14 or 28 d), antenna number (14 or 28 d), and sexual maturation (28 d); *C. riparius* survival; *D. magna* reproduction; *L. minor* frond growth; and *C. dubia* reproduction. Group 2 included *C. dubia* survival, *C. riparius* length, *H. azteca* sexual maturation (14 d), and *P. promelas* larval weight. Three endpoints comprised the third grouping, including *H. verticillata* root length, *Diporeia* (*P. hoyi*) preference, and *H. azteca* survival (14 d). *Hexagenia bilineata* survival and molting explained the fourth pattern of responses. All of the endpoints included in these groupings were whole sediment exposures.

TABLE 4. Principal component analysis of ARCS sites.

Assay	Factor[a] 1	2	3	4
Chironomus riparius 14 d survival	.97			
Chironomus tentans 10 d length	.96			
Hyalella azteca 28 d antenna length	.95			
H. azteca 28 d length	.94			
H. azteca 14 d antenna length	.94			
Daphnia magna 7 d reproduction	.92			
Lemna minor 4 d frond	.90			
H. azteca 14 d length	.86			
H. azteca 28 d survival	.63			
Ceriodaphnia dubia 7 d reproduction	–.74			
H. azteca 28 d sexual maturation	–.94			
C. dubia 7d survival		.85		
C. riparius 14 d length		.83		
H. azteca 14 d sexual maturation		–.70		
Pimephales promelas 7 d larval weight		–.71		
Hydrilla verticillata 10 d root length			.92	
Diporeia 5 d preference			.78	
H. azteca 14 d survival			–.60	
Hexagenia bilineata 10 d survival				.91
H. bilineata 10 d molting frequency				.72

These findings suggest that responses within each group are producing similar and redundant information. If a test battery were to be selected that detected each type of toxicity response pattern (Factors 1 to 4), one assay consisting of two or more endpoints could provide unique information for multiple groupings. For example, the *H. azteca* 14 d assay consisting of survival, length, antenna number, and sexual maturation endpoints is representative of three unique response patterns, while only *H. bilineata* describes the fourth pattern. Both the *C. dubia* and *C. riparius* assays can be used to explain factors 1 and 2. Use of these assays would enable each unique response pattern to be covered with fewer organism types.

Correlations Between Assay Responses

Correlating the endpoint responses (both laboratory assays and community structure) to detect similar response patterns is another useful method to evaluate data redundancy and provide field validation of toxicity assays. All measured endpoints were correlated with each other and the top ten correlations evaluated based on the r^2 and p values.

The numbers of significant correlations between assay responses vary with the degree of site contamination. Indiana Harbor was the worst contaminated (Nelson *et al.* 1996) and most toxic site of the three surveyed. Indiana Harbor had the highest number of significant (greater than or equal to p = 0.05) correlations. The Buffalo River samples exhibited less contamination and toxicity compared to the other sites and had the fewest correlations. The Saginaw River No. 1 survey had a moderate level of toxicity. There was little toxicity observed in the Saginaw 3 sediments.

A review of the best 10 correlations for each assay (93 endpoints) at each site showed 72% of the endpoints had more than 10 significant correlations and 77% had endpoint correlations with r^2 greater than 0.80. Endpoints with the fewest correlations included *H. verticillata* root and shoot growth (no significant correlations), *L. minor* biomass and benthic taxa richness (two correlations), percent flatworms and microbial galactosidase activity (three correlations), *D. magna* 7 d survival (four correlations), and *L. minor* chlorophyll *a* (five correlations). The response patterns among assays was similar for sediments collected from the Indiana Harbor and Saginaw No. 1 surveys. Since there were only three samples collected in the Saginaw No. 1 survey, correlations would be similar, particularly since one

sample (station no. 6) was very toxic. The endpoints with the highest average correlation (r^2 value) were (in rank order) Microtox, *C. tentans* length, and % chironomids and % tolerant species in the artificial substrate samples. Fewer significant correlations were noted a second survey of the Saginaw River that had more samples and less toxicity associated with the samples. The high number of significant correlations between laboratory toxicity assay responses and some artificial substrate benthic macroinvertebrate endpoints (e.g., % tolerant species and % chironomids) provides a high degree of field validation for the laboratory assays.

When assessing sediment toxicity it is important to consider effects on both benthic and nonbenthic species, since there may be interactions between sediment and overlying water compartments and between benthic and nonbenthic species. Of the nonbenthic species, the *P. promelas* and cladoceran assays are the most commonly used in sediment testing. Fish and cladocerans feed on the sediment surface during whole sediment exposures, which increases their exposure. When toxicity response patterns were compared between benthic and non-benthic species there were many significant correlations. The 7-d assays with *C. dubia, D. magna*, and *P. promelas* larval growth were significantly correlated with 10 to 70% of the benthic responses. The endpoint responses of *H. azteca* were significantly correlated with 10 to 80% of the nonbenthic responses, and *C. tentans* and *C. riparius* responses were correlated with greater than 60 and 70%, respectively, of the nonbenthic responses. The indigenous sediment microbial enzyme activities correlated with 10 to 70% of the nonbenthic endpoints.

For non-dredged material applications it is preferable to conduct liquid-phase toxicity tests with interstitial water. Toxicity tests with interstitial water are better than elutriates for evaluating the potential *in situ* effects of contaminated sediment on aquatic organisms (Ankley *et al.* 1991). Elutriate tests are most appropriately used in dredged material evaluations because they mimic disposal conditions. Elutriate samples are typically less toxic than either whole-sediment or interstitial water samples (Sasson-Brickson and Burton 1991, Ankley *et al.* 1991).

CONCLUSIONS

The process of selecting the optimal assay(s) for use in an ecosystem assessment is not simple or straight forward. The optimal assay can only be selected when the objectives of the study and associated data quality objectives have been defined and there is a reasonable understanding of the physical, chemical, and biological characteristics of the study site.

A number of useful assays have been evaluated in freshwater and marine studies (Burgess and Scott 1992, Burton 1991, Lamberson *et al.* 1992, Burton and Scott 1992). It is apparent from these previous studies and the current one that no one assay, indicator, or measured response (endpoint) is superior to all others. The rankings of sensitivity and discriminatory ability varied between test sites. However, the overall rankings give an indication of which assays tend to be the most useful. To reduce uncertainty and reduce the chance of obtaining false positive or false negative results, it is important to test more than one species. The importance of testing multiple species increases with the importance of protecting the ecosystem and the need to define "significant" contamination in the "gray" (marginally contaminated) zone.

Each assay provides information that is unique to that species and the life process measured (e.g., survival, growth). Some of assay responses are similar to each other, thus provide duplicative (redundant) information. If this similarity between assay response patterns occurs at several test sites, then some of the similar assays could be deleted from a test battery, using other criteria for assay selection, e.g., cost, resource requirements, difficulty. Resources might be put to better use by deleting assays that produce redundant information and measuring unique toxicant response patterns, such as observed in the PCA analyses (Table 4). However, if one wants to take a "weight-of-evidence" approach, use of assays which respond similarly to contaminated sediments may provide additional effect information. This increases the uncertainty, however, that a realistic comprehensive assessment of ecosystem effects is being attained. A toxicity survey which is too limited in scope may produce false positive or false negative results and thereby faulty conclusions. It may be more important to conduct fewer assays with more samples to better define the spatial and temporal variability in sediment contamination.

Other critical factors to consider in selection of the test responses are relative abilities at detecting sediment toxicity (i.e., sensitivity) and measuring level of toxicity (i.e., discrimination). Sediment toxicity appeared to relate well to the relative degree of chemical contamination at these sites. To relate these assay responses to chemical contaminant levels requires detailed analyses (described in In-

gersoll *et al.* 1996, and USEPA 1994b). Sediment contamination can vary dramatically within and between sites. Additional analyses are required to determine if the endpoints listed in Table 1 are responding to contaminants associated with sediments or to physical (e.g., grain size) or non-contaminant chemical characteristics, (e.g., interstitial water hardness). An assay that was sensitive and discriminatory at one site will not necessarily respond in a similar fashion at another site.

Integrative studies should use a water column and benthic species in whole-sediment exposures as resources permit. Useful water column assays are 7-d early life stage exposures with fathead minnows (*Pimephales promelas* larval growth or embryo-larval growth and terata (abnormalities) and 7-d survival) and reproduction assays with *Ceriodaphnia dubia* or *Daphnia dubia*. Both of these cladoceran assays were recently developed as ASTM methods (ASTM 1995). Benthic assays that were most successful were: (1) 7- to 28-d survival, growth, and development test with *Hyalella azteca*, (2) 10- to 14-d survival and growth assays with *Chironomus tentans* or *Chironomus riparius*, and (3) the 10-d survival and molting assay with *Hexagenia bilineata*. Methods for testing sediment with the amphipod *Diporeia* spp. (formerly *Pontoporeia hoyi*) have been recently described (ASTM 1995). Survival and sediment avoidance or preference of *Diporeia* are monitored in 28 and 5 d exposures, respectively (Burton *et al.* 1989).

The choice of the appropriate endpoint (response) to measure is important to the assessment process. All toxicants do not affect the same metabolic processes and result in the same effects since they have differing modes of action and target receptors. Some toxicants may interfere with processes essential for reproduction or growth. Relative species sensitivity frequently varies among contaminants. For example, Reish (1988) reported the relative toxicity of six metals (As, Cd, Cr, Cu, Hg, and Zn) to crustaceans, polychaetes, pelecypods, and fishes and concluded that no one species or group of animals was the most sensitive to all of the metals. Differences in contaminant types and concentrations between sites likely contributes to differences in species responses; however, the present study sites were each contaminated with a wide range of PAHs and metals.

Contaminants may also stimulate a process due to interruption of a feed-back mechanism or they may be essential nutrients at low concentrations (e.g., selenium). Stimulation at low concentrations of toxicant exposure (hormesis) is often reported in the literature. Some responses are much more sensitive than others (e.g., enzyme inhibition vs. lethality), and should not necessarily be weighted equally in evaluating the importance of effects. Therefore, a battery of tests are required to identify toxic sediment samples.

A critical issue that is not directly addressed with laboratory toxicity tests is the ecological significance of the measured endpoints. As discussed above, there were many significant correlations between laboratory responses and benthic community structure patterns. The relationship to chemical contaminant levels was less distinct, possibly due to differing bioavailability levels. However, general trends were evident between extremely contaminated or slightly contaminated sites and test species effects. The most sensitive toxicity endpoint in the ARCS project was the avoidance or preference behavior of *Diporeia*, a common amphipod in the Great Lakes. Behavior is often a sensitive indicator of sublethal responses. What is not known is whether or not the preference for one sediment over another would alter the population, community, or ecosystem to any degree that constitutes short- or long-term impairment. These issues are best resolved by a "weight-of-evidence" approach whereby other toxicity endpoints and community analyses are considered along with chemical and physical characteristics.

For most applications, a minimum test battery consisting of two to three assays should be evaluated. These recommendations are for waters in the United States and are based on the above characteristics and on comparison studies where multiple species have been used simultaneously in sediment contamination investigations (Burton 1991, Burton *et al.* 1989, Burton and Scott 1992, Giesy *et al.* 1988, Giesy and Hoke 1990, Hoke *et al.* 1990, Ingersoll *et al.* 1993, Chapman *et al.* 1992, Long and Buchman 1989).

It appears from these data that several measured endpoints would be useful for routine sediment contamination assessments. Results from the statistical analyses indicate two test species (with four measured endpoints) could be used to describe the three major toxicity response patterns observed at the ARCS test sites (Table 4). The endpoints that could be selected vary in their sensitivity, discrimination of toxicity, relationship to other assay responses and benthic community indices, and other strengths and weaknesses.

The following recommendations for selection of

optimal assays in future assessment of contaminated Great Lakes sediment are based on sensitivity, discrimination, and similarity analyses and on their strengths and weaknesses associated with each method. It is evident that the optimal assays vary between sites and this variation cannot be confirmed *a priori*. The PCA analysis provides an approach for selection of assays to be included in a test battery. Species can be chosen with endpoints representing each of the major response pattern groupings identified in Table 4, to better ensure that the many varied and potentially adverse species responses are being evaluated. Many of the assays that appeared best in the PCA analysis and in the sensitivity and discriminatory analyses have also been demonstrated to be good indicators of sediment toxicity in previous assessments (Burton 1991). Based on the current study, the minimal test battery recommended for Great Lakes sediment toxicity studies should consist of two species, four measurement endpoints, and represent three of the four major response pattern groups (Table 5). This enables some flexibility in the choice of the test species which may be based on other descision criteria, such as resource requirements, laboratory expertise or organism availability, need for sensitivity or discriminatory power, or other characteristics.

The first test battery option consists of two species. The only assay whose endpoints characterized three of the four response patterns was the *H. azteca* 14 d test, consisting of survival, length, and sexual maturation. Unfortunately, measuring amphipod length accurately requires use of digitizing microscope equipment. Perhaps dry weight could be measured instead of length. Furthermore, antennal segment number is a good predictor of *H. azteca* length. While 14-d tests were conducted with *C. riparius* and *H. azteca*, the USEPA (1994a) described methods for testing amphipods and midges in 10-d exposures, monitoring survival and growth. In combination with this amphipod, any of five different assays should be tested, including: *C. dubia* 7-d survival and reproduction, or *C. riparius* 14-d survival and length, or *P. promelas* 7-d larval growth, or *Diporeia* 5-d preference, or *H. bilineata* 10-d survival and molting assay.

The second test battery option consists of two species (Table 5). The first is either *C. dubia* or *C. riparius*. The second species should be *Diporeia* or *H. bilineata*. The third option for a test battery should consist of three species (Table 5). The first is *D. magna*, the second is *P. promelas*, and the third is either *Diporeia* or *H. bilineata*.

There is some degree of confidence in the conclusions reached from the large ARCS toxicity data set (7,600 data points), that organisms that were relatively sensitive or discriminatory at three or four of the sites in the ARCS project would probably be sensitive or discriminatory at other sites. It should be noted that those assays which are recommended based on the ARCS study are similar to those recommended in other North American studies by the International Joint Commission (1988), Giesy and Hoke (1990), Giesy *et al.* (1988, 1989), Kemble *et al.* (1994), Burton *et al.* (1989), USEPA (1994), and ASTM (1995). There have been no studies comparing sediment toxicity test methods that were as comprehensive as the present study.

Additional assays could be used in these test batteries, and additional measured responses will strengthen the weight-of-evidence, reducing uncertainty. Many of the assays that ranked high in the ARCS project have been used successfully in other studies of sediment toxicity. The *C. tentans* 10-d assay was not included in the minimal test batteries, due to data limitations in the discriminatory analyses. However, the growth endpoint was shown to be

TABLE 5. Optimal groupings for test battery.[1]

Group 1:	a. *Hyalella azteca* 14 d survival, length and sexual maturation
	b. *Ceriodaphnia dubia* 7 d survival and reproduction, or *Chironomus riparius* 14 d survival and length, or *Daphnia magna* 7 d survival and reproduction, or *Pimephales promelas* 7 d larval survival and weight, or *Diporeia (Pontoporeia hoyi)* 5 d avoidance/ preference, or *Hexagenia bilineata* 10 d survival and molting frequency.
Group 2:	a. *C. dubia* or *C. riparius* (endpoints listed above)
	b. *Diporeia* or *H. bilineata*
Group 3:	a. *D. magna*
	b. *P. promelas*
	c. *Diporeia* or *H. bilineata*
Reconnaissance Surveys: Microtox	

[1]Based on Principal components analysis, sensitivity, discriminatory ability, and correlation with other measurement endpoints.

a sensitive response measure in this and other studies and should be considered as useful assay. In addition, data requirements prevented principal component analysis of some useful assays such as *D. magna* and *P. promelas* 7-d assays. So a variety of toxicity assays exists, representing differing biological levels, which are sensitive and discriminatory. Evaluations of sediment using laboratory toxicity tests and benthic community structure indices, combined with physico-chemical characterization of the test site will allow for an integrated "weight-of-evidence" assessment approach which can be used to provide evidence of contaminant-induced degradation to aquatic communities.

ACKNOWLEDGMENTS

The current study was funded by U.S. Environmental Protection Agency Contract R005907-01.

REFERENCES

American Society for Testing and Materials. 1995. Standard test methods for measuring the toxicity of sediment-associated contaminants with freshwater invertebrates. E1706–95a. In *Annual Book of Standards*, Vol. 11.05, Philadelphia, PA.

Ankley, G.T., Schubauer-Berigan, M.K., and Dierkes, J.R. 1991. Predicting the toxicity of bulk sediments to aquatic organisms with aqueous test fractions: pore water vs. elutriate. *Environ. Toxicol. Chem.* 10: 1359–1366.

_____, Schubauer-Berigan, M.K., and Dierkes, J.R. 1996. Application of toxicity identification evaluation techniques to pore water from Buffalo River sediments. *J. Great Lakes Res.* 22:534–544.

Burgess, R.M., and Scott, K.J. 1992. The significance of in-place contaminated marine sediments on the water column: processes and effects. In *Sediment Toxicity Assessment*, G.A. Burton, Jr. (ed.), pp. 129–166. Boca Raton, FL: Lewis Publishers.

Burton, G.A., Jr. 1991. Assessing freshwater sediment toxicity. *Environ. Toxicol. Chem.* 10:1585–1627.

_____. 1992. *Sediment Toxicity Assessment.* Boca Raton, FL: Lewis Publishers.

_____, and Scott, K. J. 1992. Sediment toxicity evaluations: Their niche in ecological assessments. *Environ. Sci. Technol.* 26: 2068–2075.

_____, Stemmer, B.L., Winks, K.L., Ross, P.E., and Burnett, L.C. 1989. A multitrophic level evaluation of sediment toxicity in Waukegan and Indiana harbors. *Environ. Toxicol. Chem.* 8:1057–1066.

Chapman, P.M., Power, E.A., and Burton, Jr., G.A. 1992. Integrative assessments in aquatic ecosystems. In *Sediment Toxicity Assessment,* G.A. Burton, Jr., ed., Boca Raton, FL: Lewis Publishers.

Giesy, J.P., and Hoke, R.A. 1990. Freshwater sediment quality criteria toxicity assessment. In *Sediments: Chemistry and Toxicity of In-Place Pollutants*, R. Baudo, J, Giesy, and H. Mantau, eds., pp. 265–348. Boca Raton, FL: Lewis Publishers.

_____, Graney, R.L., Newsted, J.L., Rosiu, C.J., Benda, A., Kreis, Jr., R.G., and Horvath, F.J. 1988. Comparison of three sediment bioassay methods using DetroitRiver sediments. *Environ. Toxicol. Chem.* 7:483–498.

_____, Rossiu, C.J., Graney, R.L., and Henry, M.G. 1989. Benthic invertebrate bioassay with toxic sediment and pore water. *Environ. Toxicol. Chem.* Vol. 9, 233–248.

Gossiaux, D.C., Landrum, P.F., and Tsymbal, V.N. 1993. A survey of Saginaw River and Saginaw Bay, Lake Huron, sediments using two bioassaays with the amphipod *Diporeia* spp. *J. Great Lakes Res.* 22:322–332.

Hall, N.E., Fairchild, J.F., La Point, T.W., Heine, P.R., Ruessler, D.S., and Ingersoll, C.G. 1996. Problems and recommendations in using algal toxicity testing to evaluate contaminated sediments. *J. Great Lakes Res.* 22:545–556.

Hoke, R.A., Geisy, J.P., Ankley, G.T., Newsted, J.L., and Adams, J.R. 1990. Toxicity of sediments from Western Lake Erie and the Maumee River at Toledo, Ohio, 1987: implications for dredged material disposal practices. *J. Great Lakes Res.* 16:457–470.

_____, Buckler, D.R., Crecelius, E.A., and La Point, T.W. 1993. *NRCRC–Battelle Final Report for the USEPA GLNPO Assessment and Remediation of Contaminated Sediments (ARCS) Project: Biological Assessment of Contaminated Great Lakes Sediment.* EPA.

Ingersoll, C.G., Buckler, D.R., Crecelius, E.A., and La Point, T.W. 1993. *NRCRC-Battelle Final Report for the USEPA GLNPO Assessment and Remediation of Contaminated Sediments (ARCS) Project: Biological and Chemical Assessment of Contaminated Great Lakes Sediment.* U.S. Environmental Protection Agency, Great Lakes National Program Office, Chicago, IL.

_____, Haverland, P.S., Brunson, E.L., Canfield, T.J., Dwyer, F.J., Henke, C.E., and Kemble, N.E. 1996. Calculation and evaluation of sediment effect concentrations for the amphipod *Hyalella azteca* and the midge *Chironomus riparius. J. Great Lakes Res.* 22:602–623.

International Joint Commission. 1988. *Procedures for the Assessment of Contaminated Sediment Problems in the Great Lakes.* IJC. Windsor, Ontario.

Kemble, N.E., Brumbaugh, W.G., Brunson, E.L., Dwyer, F.J., Fairchild, J.F., Monda, D.P., and Woodward, D.F. 1994. Toxicity of metal-contaminated sediments from the upper Clark Fork River, Montana to aquatic

invertebrates in laboratory exposures. *Environ. Toxicol. Chem.* 13: 1985– 1997.

Kemp, P.F., and Swartz, R.C. 1988. Acute toxicity of interstitial and particle-bound cadmium to a marine infaunal amphipod. *Marine Environ. Res.* 26:135–153.

Lamberson, J.O., DeWitt, T.H., and Swartz, R.C. 1992. Assessment of sediment toxicity to marine benthos. In *Sediment Toxicity Assessment*, G.A. Burton, Jr., (ed.),,pp. 183–212. Boca Raton, FL: Lewis Publishers.

Long, E.R., and Buchman, M.F. 1989. *An evaluation of candidate measures of biological effects for the National Status and Trends Program.* NOAA Technical Memorandum, NOS OMA 45. Seattle, WA.

Reish, D.J. 1988. The use of toxicity testing in marine environmental research. In *Marine Organisms as Indicators*, Chapter 10, edited by D.F. Soule and G.S. Kleppel, pp–231–245. New York: Springer-Verlag.

Samoiloff, M.R., Schultz, S., Jordan, Y., Denich, K., and Arnott, E. 1980. A rapid simple long- term toxicity assay for aquatic contaminants using the nematode *Panagrellus redivivus. Can. J. Fish. Aquat. Sci.* 37:1167–1174.

Sasson-Brickson, G., and Burton, Jr., G.A. 1991. *In situ* toxicity testing with *Ceriodaphnia dubia. Environ. Toxicol. Chem.* 10:201–207.

Snell, T.W., and Persoone, G. 1989. Acute toxicity bioassays using rotifers. I. A test for brackish and marine environments with *Brachionus plicatilis. Aquatic Toxicol.* 14:65–80.

Southerland, E., Kravitz, M., and Wall, T. 1992. Management framework for contaminated sediments (The U.S. EPA Sediment Management Strategy). In *Sediment Toxicity Assessment*, G.A. Burton, Jr., (ed.), pp. 341–370. Boca Raton, FL: Lewis Publishers.

U.S. Environmental Protection Agency. 1994a. *Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates.* EPA 600-R24-024, Duluth, MN.

_____. 1994b. *Assessment guidance document.* Assessment and Remediation of Contaminated Sediments Program. Great Lakes National Program Office, EPA 905-B94-002, Chicago, IL.

U.S. Environmental Protection Agency-U.S. Army Corps of Engineers. 1994. *Evaluation of dredged material proposed for discharge in inland and near coastal waters* (draft). EPA-823-B-94-002, Washington, DC.

Submitted: 15 April 1996
Accepted: 17 June 1996