An international comparison of sediment toxicity tests in the North Sea

Peter M. Chapman[1], Richard C. Swartz[2], Brian Roddie[3], Harriette L. Phelps[4], Peter van den Hurk[5], Ray Butler[6]

[1] EVS Consultants, 195 Pemberton Avenue, North Vancouver, British Columbia, Canada V7P 2R4
[2] U.S. EPA/ERL, Hatfield Marine Science Center, Newport, Oregon 97365, USA
[3] Environment and Resource Technology, Heriot-Watt University, Riccarton, Edinburgh EH14 4AS, United Kingdom
[4] Biology Department, University of Washington, 4200 Connecticut Avenue N.W., Washington, D.C. 20008, USA
[5] Bureau Waardenburg bv, PO Box 365, 4100 AJ Culemborg, The Netherlands
[6] Water Research Center, Henley Road, Medmenham, PO Box 16, Marlow, Buckinghamshire SL7 2HD, United Kingdom

ABSTRACT: Toxicity testing of whole sediments was conducted as part of the Bremerhaven Workshop designed to test various methods (chemical and biological) for assessing the status of North Sea waters, sediments and biota. Six investigators from 4 countries were involved; laboratory testing was conducted after transporting field-collected sediments distances varying from tens of miles to thousands of miles. Sediments were tested from 2 contamination gradients, one from an abandoned drilling site, and the other from the mouth of the Elbe northwest across the German Bight. Methods included 11 different tests (20 end-points), 3 species of amphipod, a polychaete, a clam, an oyster and a bacterium. Amphipod 10 d acute lethality tests and a 48 h oyster larvae abnormal development test most clearly determined gradients in toxicity that corresponded with chemical and *in situ* community data. Lack of response was observed in Microtox and clam reburial tests. A polychaete growth test conducted in North America provided useful but not convincing information. A 24 h oyster larvae survival test conducted separately in England and The Netherlands gave results that were counter to the other tests and difficult to interpret. Survival and metamorphosis tests with older oyster larvae did not show consistent, interpretable gradients for the drilling site but did for the German Bight. The responsiveness of some tests may have been affected by a maximum 3 wk sediment holding time prior to testing; other tests did not provide usable information relative to either the burden of evidence of all tests, or corresponding chemical contamination data and benthic infaunal community structure. Based on the results of this workshop, currently the most useful sediment toxicity tests for general assessment and regulatory use in Europe are infaunal amphipod survival and 48 h oyster larval development tests.

INTRODUCTION

Sediment toxicity tests involving the exposure of laboratory organisms to field-collected sediments, either directly (whole sediment tests) or indirectly (via elutriates or extracts of the sediments) are becoming increasingly important in regulation and assessment in North America (Chapman 1986, 1988, Chapman et al. 1991, Thompson 1991). However, such tests have not yet been used in regulation in Europe and are, in fact, generally still in various stages of development.

In March 1990, as part of the Bremerhaven Workshop, sediment toxicity tests developed in North America and Europe were applied to 2 transects across expected gradients of chemical contamination in the North Sea (Fig. 1). One transect comprised 7 stations at increasing distances down the residual current from an abandoned drilling site off the Dutch coast (see Daan et al. 1992). The other transect comprised 9 stations at increasing distances from the mouth of the Elbe River, Germany, out to the Dogger Bank (see Stebbing & Dethlefsen 1992). The overall aim of the Workshop was to compare biological effects monitoring techniques for marine pollution, and to determine the most suitable techniques for such monitoring in Europe.
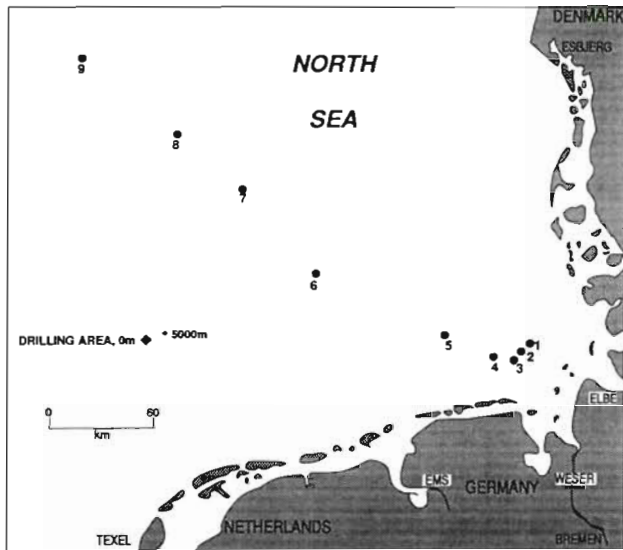
Fig. 1. Location of sampling stations. Stns 1 to 9 are German Bight stations. Drilling Area stations are 0, 125, 250, 500, 1000, 2000 and 5000 m from the center of drilling

The purpose of this paper is to: (1) report on the results of the overall sediment toxicity testing, and (2) recommend suitable sediment toxicity test techniques for

Europe. Details of the tests conducted are provided in Table 1. Separate papers with more detail on specific aspects of the testing are provided by van den Hurk et al. (1992) and Butler et al. (1992).

METHODS

Three replicate sediment samples were collected from each station at the drilling site on March 13 to 15, 1990. Collection was originally by a 0.13 m^2 Reineck box corer but, due to adverse weather conditions, sediments from the 125 and 250 m stations (Stns B & C) were collected with a 0.1 m^2 Van Veen grab. A scooped portion of the top 20 cm of sediment was removed intact from the corer or grab using precleaned (acetone and seawater) stainless steel instruments, placed in clean, labelled polyethylene bags and kept cool (about 6 °C); the remainder was used for chemical analysis (see Cofino et al. 1992). On March 17, the samples were removed from the bags, homogenised in precleaned stainless steel bowls (Chapman 1988), then split into separate subsamples for the 5 different testing laboratories. Subsamples were placed in clean, labelled polyethylene bags from

Table 1. Marine sediment tests conducted

Investigator	Country	Test organism	End point	Length of test	Route of exposure
Swartz	USA	Rhepoxynius abronius (amphipod)	Survival Reburial	10 d	Whole sediment
Chapman	Canada	Rhepoxynius abronius (amphipod)	Survival Avoidance Reburial	10 d	Whole sediment
		Neanthes arenaceodentata (polychaete)	Survival Growth	20 d	Whole sediment
		Crassostrea gigas (oyster)	Survival Development	2 d	Elutriate
Roddie/Butler	UK	Corophium volutator (amphipod)	Survival Immobilisation	10 d	Whole sediment
		Crassostrea gigas (oyster)	Survival Development	24 h	Elutriate
		Microtox (bacteria)	Bioluminescence	< 1 h	Whole sediment
Phelps	USA	Mya arenaria (clam)	Burrowing	~2 d	Whole sediment
		Crassostrea gigas (oyster)	Survival Metamorphosis	~2 d	Whole sediment
van den Hurk	The Netherlands	Crassostrea gigas (oyster)	Survival	24 h	Elutriate
		Bathyporeia sarsi (amphipod)	Survival Reburial	10 d	Whole sediment

which air had been excluded, and were shipped in coolers on ice.

Replicate sediment samples were collected from the 9 German Bight stations on March 27 to 29, 1990 using Van Veen grabs. For reasons associated with permission to enter territorial waters, sampling was conducted from 2 separate vessels; 4 replicates of the top 4 cm of sediment were collected by 4 separate 0.1 m^2 Van Veen grabs from the 6 innermost stations (Stns 1 to 6) and 3 replicates of the top 10 cm of sediment were collected by 0.2 m^2 Van Veen grab from the 3 outermost stations (Stns 6 to 9). Stn 6 was sampled by both vessels; these duplicate samples were tested by all but 2 tests: oyster larvae testing by Phelps and amphipod testing by van den Hurk. The inadvertent provision of overlapping samples was fortuitous due to differences in sampling between the 2 vessels. Sample storage and splitting were handled after collection similarly to those samples from the drilling site; homogenisation and splitting were done on March 29. The only difference was that chemistry samples comprised post-collection subsamples of the homogenised sediment allocated to each laboratory.

All toxicity tests were initiated within 4 wk of sediment collection; most tests were initiated within 2 to 3 wk. All test organisms were properly acclimated prior to testing. Sand, silt, clay and loss on ignition (LOI) were determined at bioassay Day 0 (Swartz, amphipod tests) on a composite sample of all replicates at each station.

Quality assurance/quality control (QA/QC) of toxicity testing generally included, as per Chapman (1988): use of positive and negative controls, use of healthy test organisms, confirmed taxonomic identifications, standard laboratory procedures, randomised (blind testing where possible) container placement, and maintenance and measurement of water quality conditions. Positive controls were not used for the Dutch oyster larvae test or the U.S. oyster survival and metamorphosis test.

All 3 amphipod species (*Rhepoxynius abronius*, *Corophium volutator* and *Bathyporeia sarsi*) were exposed to solid phase sediments using the procedures of Swartz et al. (1985) and Chapman & Becker (1986). Exposure was in 1 l glass beakers containing a 2 cm layer of test sediment overlaid with filtered clean seawater. Constant aeration of the overlying water, and requisite environmental conditions were provided. *C. volutator* tests were performed with 10 individuals per beaker under 16:8 h light:dark conditions; the other species were tested with 20 animals per beaker. *R. abronius* was tested under constant light conditions; *B. sarsi* was tested under 16:8 h light:dark conditions. After a 10 d exposure period the sediments were sieved and surviving amphipods removed and

counted. Ancillary information for particular tests included avoidance, reburial and immobilisation (Table 1). Dilution tests (50 %, 10 %, 1 %) were conducted with *R. abronius* by EPA with Stn 1, 2 & 3 sediments based on an expectation of toxicity at these stations. Laboratory replication was conducted for some tests, but not all; further details of amphipod testing are provided in van den Hurk et al. (1992).

Neanthes arenaceodentata solid phase sediment tests followed the protocol of Johns et al. (1990) and were the same as the amphipod tests with the following exceptions. Five juvenile worms were exposed to each replicate test sediment in a static renewal exposure system (every third day, one-third of the water was replaced). Two laboratory replicates were tested for each field replicate of the drilling site sediments; there was no laboratory replication of the German Bight sediments. During the 20 d exposure period, worms were fed 40 mg of crushed Tetra Marin flakes (i.e. 8 mg per individual) on an every-other-day basis. Prior to test initiation, initial worm biomass mg dry weight) was determined on 3 subsamples of 5 worms. At test termination, the sediments were sieved and surviving worms removed and weighed to determine final biomass. Control sediment was from a collection site for *R. abronius*: West Beach, Whidbey Island, Washington, USA. Final calculations included estimated individual dry weights and growth rate.

Bivalve larvae bioassays in all cases used the Pacific oyster *Crassostrea gigas*. However, different techniques were used by different laboratories for the egg to larval tests. These are fully described by Butler et al. (1992) and are summarised below.

Testing in Canada was conducted for 48 h following the procedures of ASTM (1986) and Chapman & Becker (1986). Adult oysters were induced to spawn by thermal and biological stimulation. Eggs were washed through a 250 µm Nitex screen to remove excess gonadal material. Fertilisation was accomplished by combining all the eggs from one female and 20 to 50 ml of sperm from a male oyster in a 2 l beaker. Embryos were kept suspended by gentle agitation with a perforated plunger. Embryo density was determined by duplicate counts of a 1:99 dilution of the embryo suspension. Testing was conducted in pre-cleaned 1 l polyethylene jars containing 20 g of sediment and filled with clean seawater. Sediments were shaken for 10 s and allowed to settle for 1 h before inoculation. Each container was inoculated with 23 000 embryos within 2 h of fertilisation (stocking density was determined from initial counts using sacrificial jars). Two laboratory replicates were tested for each field replicate of the drilling site sediments; there was no replication of the German Bight sediments. After a 48 ± 4 h

exposure period under a 14:10 h light:dark photoperiod without aeration, the test was terminated by the direct pipette method. Representative samples were preserved and examined in a Sedgewick-Rafter counting chamber at 100× magnification. Survival was determined. Normal and abnormal prodissoconch I larvae were enumerated to determine percent abnormality (= failure to transform to the fully shelled, straight-hinged, 'D-shaped' larva).

Testing in the UK differed from the Canadian methodology in the following particulars. Elutriate was prepared by mixing 0.25 l of sediment with 0.25 l of reference seawater in 0.5 l polyethylene containers; the containers were rolled mechanically for 16 h, allowed to settle for 1 h, then supernatant was decanted and filtered through 0.45 μm membrane filters. A range of sediment extract dilutions was tested in triplicate: 1, 10, and 100 % for both areas; 30 % and 50 % were also tested for, respectively, the drilling site and the German Bight. Density of oyster embryos in test containers was similar to the Canadian technique: 20 000 l⁻¹. A major difference was the fact incubation was in the dark for 24 h at 25 °C as opposed to 48 h at 20 °C under a light:dark regime. Termination of the test was by preservation of the whole sample then removal of aliquots for counting. Numbers of normal D-shaped larvae were expressed as percentages of total survivors.

Testing of oyster larvae in The Netherlands combined North American (Chapman & Morgan 1983) and English (Thain 1991) procedures. Major differences were that oysters were stripped rather than induced to spawn, fertilised eggs from different females were pooled, density of oyster embryos was almost 3× higher (53 000 l⁻¹), and surviving larvae were counted electronically (by Coulter Counter) rather than by visual determination of normal development. Exposure was 24 h at 24 °C under ambient light. Elutriate was prepared similarly to the Canadian method but with a longer period of hand shaking. Termination of the test was by siphoning approximately 80 % of the overlying water through a 30 μm filter, preserving and counting the residue on the filter.

Two other bivalve tests were conducted, in the USA.: a rapid burrowing test using the clam *Mya arenaria* (Phelps 1989, 1990), and a 96 h measure of mortality and metamorphosis of *Crassostrea gigas* oysters (Phelps & Warner 1990). The burrowing bioassay involved placing 20 clams on 1 l of sediment (depth 4 to 6 cm) with 5 cm of clean overlying seawater and determining, at 5 min intervals, the number of clams that had completely reburied. An ET_{50} (time to reburial of 50 % of clams) was calculated. The oyster test involved obtaining oyster pediveliger larvae from a West Coast oyster hatchery (testing was conducted on the East Coast), and inducing onset of metamorphosis in pedi-

veliger larvae by exposing them to a solution of epinephrine. Thirty larvae were then exposed to sediment pressed through a 149 μm Nytex screen. Exposure was in 2.75 ml wells of tissue culture plates containing a layer of sediment and overlying clean artificial seawater at 21 °C. After a 96 h exposure period, sediments were filtered through a 149 μm Nytex screen and both survival and successful metamorphosis of the larvae determined. Metamorphosis was determined as transition from rounded veliger to a flat shape with new shell growth (Phelps & Warner 1990).

Microtox testing of the sediments involved direct exposure methods (Microtox Direct Assay; Tung et al. 1991) under development at the time of testing. Microtox test reagents containing the luminescent micro-organisms were mixed directly with serial 50 % dilutions of sediment suspension starting at 0.1 g ml⁻¹. After a 20 min exposure period at room temperature followed by 5 min at 15 °C, the suspended particulates (excluding the bacteria) were selectively removed by specialised filtration; light intensity was then measured and a 15 min EC_{50} calculated. Test reproducibility was evaluated by randomly retesting 4 sediment samples. Controls consisted of sediment from an area known to be uncontaminated (Poole, UK) and an area of known contamination (Plymouth, UK).

Means and 95 % confidence limits were determined. Significant differences in end-points were initially determined by ANOVA or the Kruskal-Wallis test in the case of unequal variances and, where appropriate, by a *t*-test comparison of means or Dunnett's procedure (Steel & Torrie 1960, Zar 1984).

RESULTS

Most tests met acceptable negative (clean) control criteria of less than 10 % effects (e.g. mortality, abnormality). However, UK mean oyster larvae negative control survival for the drilling site samples was slightly lower: 87 %. Dutch oyster larvae tests were much lower: 19 % for the German Bight samples and 8 % for the drilling site samples. These latter severe survival problems experienced with seawater blanks occurred during other testing by this laboratory in 1989 (van den Hurk unpubl.) and may be an artifact of the test methods (Butler et al. 1992). Water quality conditions in all tests were within acceptable limits. Positive (reference toxicant) control results were within expected limits for all tests with the exception of *Rhepoxynius abronius* exposed in Canada to drilling site sediments, which was more sensitive than expected (cf. van den Hurk et al. 1992). There were, at the time of testing, no control criteria for the direct exposure Microtox assay.

Table 2. Results of *Neanthes arenaceodentata* sediment toxicity tests

	Drilling area				German Bight		
Stn	Survival (%)	Estimated ind. dry wts (mg)[a]	Growth rate (mg d^{-1})[b]	Stn	Survival (%)	Estimated ind. dry wts (mg)[a]	Growth rate (mg d^{-1})[b]
A	100	9.18 AB	0.455 AB	1	95	12.6 AB	0.619 AB
B	100	8.90 A	0.441 A	2	80	9.6 A	0.470 A
C	100	11.5 B	0.570 B	3	100	10.3 A	0.504 A
D	96	9.42 AB	0.467 AB	4	100	15.9 AB	0.783 AB
E	100	9.47 AB	0.469 AB	5	100	15.5 AB	0.762 AB
F	96	10.0 AB	0.498 AB	6a&b	100	16.8 AB	0.829 AB
G	96	11.5 B	0.569 B	7	100	21.6 B	1.066 B
				8	100	16.7 AB	0.825 AB
				9	100	15.2 AB	0.748 AB
Control	100	7.92	0.392	Control	100	12.1	0.593

[a] Total dry biomass/number of survivors. Means with the same letter are not significantly different as determined by ANOVA

[b] $G = \dfrac{DW_t - DW_i}{T}$ where G = estimated individual growth rate (mg dry wt d^{-1}); DW_t = estimated individual dry weight at termination (mg); DW_i = mean estimated individual dry weight at initiation (mg); T = exposure time (d). Means with the same letter are not significantly different as determined by ANOVA

Means and 95 % confidence intervals of the field replicate tests are provided in Figs. 2 & 3 for, respectively, the major end-points of the amphipod and oyster larvae tests. Results for other end-points for these tests are provided in, respectively, van den Hurk et al. (1992) and Butler et al. (1992). Dilution experiments with Stn 1 & 2 sediments showed no toxicity (mean survival 93 to 98 %), which is not surprising given the relatively low toxicity of the undiluted sediments (67 to 97 %). Tables 2 & 3 provide summarised information on the *Neanthes arenaceodentata* and Microtox tests, respectively. Results of duplicate Microtox tests (for reproducibility) are shown in Table 4.

Neanthes arenaceodentata reference (positive control) 96 h LC$_{50}$ tests with cadmium chloride resulted in a value of 15 µg l^{-1}, which is within the range (12 to 22 µg l^{-1}) reported by Reish (1984) and Johns et al. (1990). Thus the worms were appropriately sensitive to the reference toxicant. The results (Table 2) showed no significant depression in growth at any station but low growth rate was observed in the controls. However, non-significant reductions in survival and growth were observed at the inshore German Bight stations (Table 2).

Similar results from duplicate toxicity testing at German Bight Stn 6 (Figs. 2 &
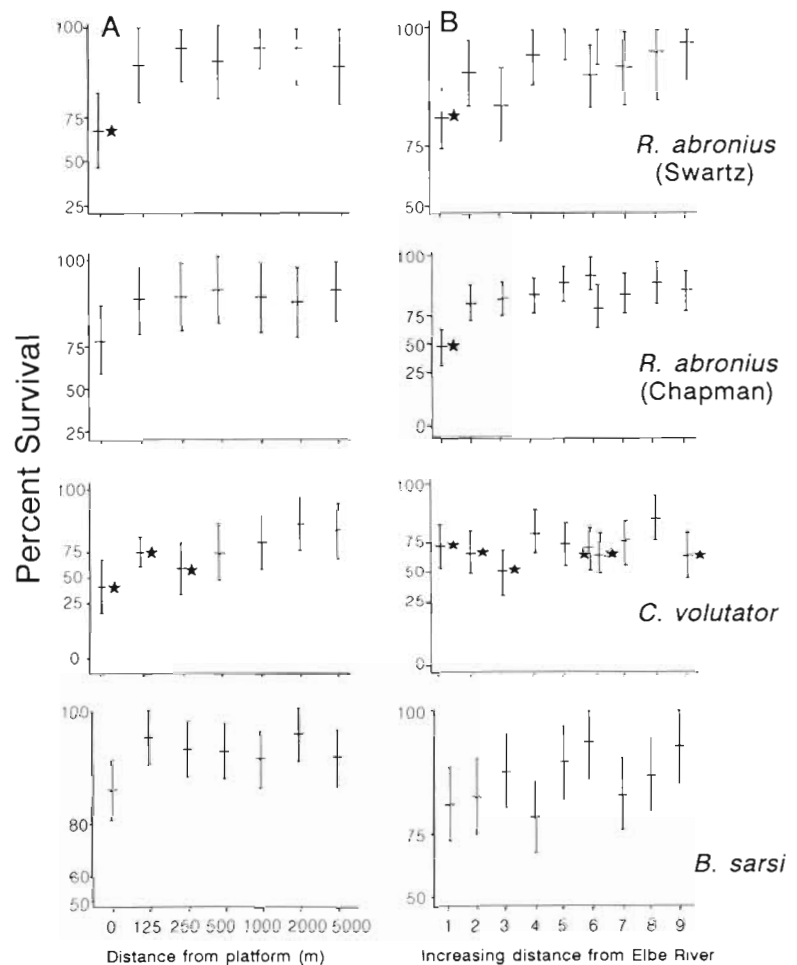

Fig. 2. Amphipod acute lethality to (A) Drilling Area sediments and (B) German Bight sediments. Means and standard deviations are shown. *: significantly different (p < 0.05) from controls
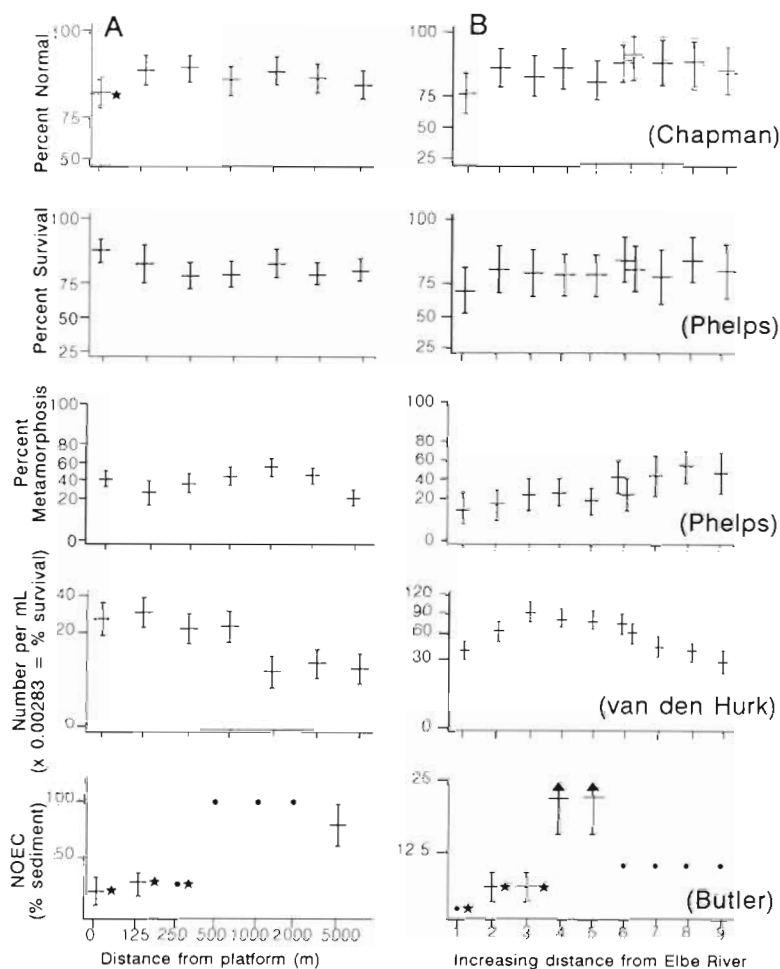
Fig. 3. *Crassostrea gigas*. Oyster larvae test responses to (A) Drilling Area sediments and (B) German Bight sediments. Means and standard deviations are shown; NOEC: no observed effect concentration. ★: significantly different (p < 0.05) from controls

3, Table 3) indicated that all German Bight stations could be treated as a single gradient despite the differences in methods of sediment collection. The alternative would have involved treating the innermost stations differently to the outermost stations.

Reburial tests with *Mya arenaria* were inconclusive, perhaps due to 2 factors. First, the test organisms appeared stressed prior to testing, which is supported by the fact that negative control (clean sand) reburial times were unacceptably long; 3 h versus the normal (Phelps 1989, 1990) 0.5 h. Second, sediments were 23 to 29 d old at the time of testing; previous work has shown that the sensitivity of the reburial response decreases dramatically with sediment age (Phelps 1989, 1990).

The direct exposure Microtox assay used in this study was concurrently under development by Microtox. The UK laboratory performing these tests did so as part of a confidential, no-cost assessment; techniques provided by Microtox were not in final form. Problems were encountered in separating particulate (and turbidity) effects from possible chemical toxicity but, despite these problems, these data appear to have provided useful information (Tables 3 & 4).

Table 5 provides grain-size data derived from bioassay samples at the time

Table 3. Results of Microtox bacterial sediment toxicity tests. Mean values; level of replication variable; see text

	Drilling site			German Bight		
Stn	15 min EC$_{50}$ (%)	95% confidence limits		Stn	15 min EC$_{50}$ (%)	95% confidence limits
A	4.8	3.2–7.7 ★		1	1.7	0.9–12.7 ★
B	6.0	4.2–8.5 ★		2	3.9	2.1–5.9 ★
C	9.2	5.1–17.1		3	4.3	2.1–9.7 ★
D	21.0	10.0–40.7		4	9.2	5.5–15.3
E	16.9	13.7–20.9		5	25.3	16.2–39.9
F	11.4	8.1–16.1		6a	18.7	12.3–29.1
				6b	21.1	16.7–26.7
G	8.2	5.8–11.9 ★		7	14.7	11.6–18.7
				8	54.6	27.3–222.3
Negative (clean) control	14.5	13.7–15.5		9	127.4	17.9–2904.3
Positive (toxic) control	2.2	0.9–6.2		Negative (clean) control	14.5	13.7–15.5
				Positive (toxic) control	2.2	0.9–6.2

★ Confidence limits do not overlap with negative (clean) control *and* do overlap with positive (toxic) control

Table 4. Microtox test reproducibility. Comparison of 4 randomly retested drilling site sediments

| Drilling site sediment | | 15 min EC_{50} (95 % confidence limits) | |
Stn	Replicate	Test	Retest
C	#3	3.4 (1.6–7.4)	3.2 (1.0–10.2)
D	#2	31.8 (11.5–88)	51.8 (16.5–>100)
E	#3	28.7 (23.4–35.1)	23.6 (18.3–30.4)
F	#1	15.7 (11.8–20.9)	20.7 (13.7–31.2)

of test initiation in one of the participating laboratories (Swartz, USA). Data provided separately by Cofino et al. (1992) as part of the sediment chemical analyses portion of the workshop indicate the same trends but measure different parameters (organic and inorganic carbon, sediment phi size). Clearly the drilling site sediments were remarkably similar while the German Bight sediments were dissimilar and showed a trend of increasing grain-size and decreasing organic carbon content with distance offshore.

DISCUSSION

Toxicity testing conducted during this study was truly international, involving laboratories in 4 countries and 2 continents. Although interlaboratory calibration and intercomparison exercises with sediment tests have been conducted previously between Canada and the U.S. (e.g. Mearns et al. 1986, Williams et al. 1986, Becker et al. 1990, Long et al. 1990), there

have been no previous comparisons involving North America and Europe.

This study was instructive not only in terms of scientific findings but also in terms of the robustness required of tests and methods which are to be used internationally and/or in large-scale studies. Despite the best efforts of all concerned, major differences occurred between the 2 transects and even within the German Bight transect during collection. These differences included different sediment samplers (corer and grab), different sizes of sediment sampler (0.1, 0.13, 0.2 m^2), different sediment depths sampled (4, 10, 20 cm), and different methods of splitting the subsamples for chemical analyses (removing an intact section of sediment from the grab versus removing a composite from homogenised sediment at the same time as composites were removed for toxicity testing). A key similarity which was preserved was the testing of field replicate samples, rather than compositing and then testing only laboratory replicates (some laboratory replication also occurred). As a result, useful data on the variability of individual sediment samples was obtained (Figs. 2 & 3, Table 3).

Overall, stations closest to sources (Stns A & 1) were generally the most toxic; however neither gradient showed a wide range of responses, nor were highly toxic (e.g. 0 % mean survival) stations among those tested. Relative degree of sediment toxicity varied from non-toxic (responses within the range of clean sediment controls) to, at worst, moderately toxic. For instance, based on relative acute toxicity data for *Rhepoxynius abronius*, 'worst case' mean survival of 46 % at Stn 1 was greater than many highly contaminated areas of the USA (mean survival 0 to 71 %; Swartz et al. 1989), but lower than the mean survival at which responses are clearly classifiable as toxic (75.5 %; Mearns et al. 1986).

Initial design of this study (e.g. prior to testing) included the following within-bioassay comparisons: between-station differences in mean response, ANOVA, multiple comparison tests and paired comparison with control responses. Between-bioassay comparisons planned included correlation coefficients between all possible pairs, using rank comparisons and/or linear regression. However, since over half the toxicity tests showed roughly the same gradient in response (effects nearest to Stns A & 1 decreasing with distance; for detailed analysis see

Table 5. Sediment grain-size and loss on ignition data

| Stn | Grain-size (%) | | | Loss on ignition (%) |
	Sand	Silt	Clay	
Drilling site				
A	84.1	9.4	6.5	1.8
B	84.4	8.3	7.4	1.7
C	84.6	8.1	7.3	1.9
D	88.9	4.8	6.3	1.8
E	94.1	3.4	2.5	1.1
F	84.3	7.0	8.7	2.2
G	85.7	8.8	5.5	1.8
German Bight				
1	6.9	57.3	35.9	10.7
2	47.1	42.6	10.3	3.7
3	49.9	31.6	18.6	7.1
4	76.7	17.7	5.6	2.3
5	78.0	20.0	2.0	1.6
6a	93.3	4.7	2.0	1.0
6b	88.5	7.7	3.8	1.1
7	89.5	5.6	4.9	1.3
8	97.0	2.1	0.9	0.5
9	100.0	0	0	0.4

Chapman 1992), simplified primarily visual comparisons were the most appropriate (cf. Figs. 2 & 3).

Drilling area macrobenthic community structure is detailed by Kröncke et al. (1992), and showed a similar trend with toxicity: decreasing number of species, biomass and total abundance within 1000 m of the platform. Although differences between stations generally were not significant, Stn A was clearly different from all other stations. For the German Bight, physical sediment variation hid any possible toxic effects although Stn 1 did have lowest number of species and abundance (Kröncke & Rachor 1992). The meiobenthos showed no evidence of pollution-related effects at any station on either gradient; this was attributed (Gee et al. 1992) to winter storms covering contaminated sediment with a layer of clean material for the drilling site gradient and, for the German Bight, to a strong riverine influence inshore coupled with grain-size differences particularly at Stn 9.

The similarity of sediment between the drilling site stations clearly indicates that sediment physical characteristics (e.g. grain-size; DeWitt et al. 1988) were not responsible for observed toxicity. Measured chemical concentrations did not match toxicity (concentrations were similar across all stations; Cofino et al. 1992). However, because not all chemicals were measured in this study, chemical toxicity cannot be excluded. Similarly, the possible effects of ammonia (Ankley et al. 1990) and AVS metal binding capacity (DiToro et al. 1990) were not addressed in this study but could account for some of the observed toxicity.

In contrast, exclusion of potential sediment physical effects is not possible for the German Bight stations, which showed a generally increasing gradient of larger grain-size and less organic matter from inshore to offshore (Table 5). This gradient matched a decreasing gradient in chemical contamination (Chapman 1992, Cofino et al. 1992) and obscured any effects that might have been detected in the macrofauna (Kröncke & Rachor 1992). It is possible that some toxicity, at least in the infaunal amphipod tests, was an artifact of natural sediment features. However, not all observed toxicity in the German Bight can be due to grain size effects: 48 h oyster larvae elutriate tests showed a similar gradient of response to the infaunal amphipods. Further, low survival in the Canadian *Rhepoxynius abronius* tests at Stn 1 (46 %) is below the 95 % Lower Prediction Limit (LPL) for survival against percent fines (DeWitt et al. 1988).

Although grain-size effects related to fine sediments have been shown to exist for *Rhepoxynius abronius* (DeWitt et al. 1988), such may not exist for some other species. For example, *Corophium volutator* has an avoidance reaction to coarse sediments, consistent with its tube-building habits and presence in tidal flats

and other habitats with muddy substrates (Bousfield 1973). Thus, this species may be ideal for testing fine (muddy) contaminated sediments. *C. volutator* may, however, be affected adversely by coarse sand such as that present at the drilling site (Meadows 1964, van de Wassenberg 1989) In this regard, it is interesting to note that *C. volutator* showed the greatest range of significant responses to the 2 gradients (3 of 7 for the drilling area; 6 of 9 for the German Bight; Fig. 2). However, differences between station replicates were equally, or more, marked (i.e. differences between replicates were in some cases as great as between stations); this result is surprising as the relative sensitivity of *C. volutator* would have been expected to be less than for *Bathyporeia sarsi* (van den Hurk et al. 1992).

Testing is presently being conducted in Canada to determine infaunal amphipod species for federal regulatory and assessment use (EVS Consultants 1991). Six species of amphipod have been tested, with emphasis on 2 of the species tested in this study: *Rhepoxynius abronius* and *Corophium volutator* (other species are *Amphiporeia virginica, Foxiphalus xiximeus, Eohaustorius estuarius* and *E. washingtonianus*). Both *R. abronius* and *C. volutator* were judged suitable for use in sediment bioassays as a result of interlaboratory calibration studies and testing across a range of contaminated sediments, although *Rhepoxynius* was judged the more suitable.

No European countries have at present recognised toxicity guidelines for sediment quality assessment, although such are in various stages of development in, for example, the Netherlands and the UK. In contrast, both the USA and Canada have such guidelines in the form of recommended tests and protocols, and in some cases testing is mandated by legislation. The results of the present study indicate that the amphipod tests being developed in Europe (*Corophium volutator* in the UK and *Bathyporeia sarsi* in The Netherlands) are appropriate for sediment toxicity determinations. Both amphipod tests showed good quality control, were able to identify contamination gradients, and conformed to the 'burden of evidence' from all tests conducted in this study. In contrast, the oyster larvae tests used by laboratories in both the UK and The Netherlands do not appear (at least in their present form and stage of development) to be appropriate for sediment toxicity determinations. Since the North American test and longer exposure method (48 vs 24 h) did provide usable and relevant results, it would be prudent to adopt these oyster larvae test methods for use in Europe.

Future development of European (and other) sediment toxicity tests should concentrate on third generation tests as defined below. Sediment toxicity tests can be divided, at present, into 3 categories. First genera-

tion tests were originally devised for water column testing and then adapted for sediment testing. The oyster larvae abnormality and Microtox tests used in this study fall into this category. All but one of the other tests (i.e. amphipod, clam, oyster metamorphosis) are second generation tests, comprising acute lethal tests devised specifically for sediments. Third generation tests are those devised specifically for sediments, which measure growth and/or reproduction in addition to survival and which thus approach chronic testing (defined as full life-cycle tests). The *Neanthes arenaceodentata* test falls into this final category, but did not prove as useful in this study as expected, because growth in the control worms was almost always less than that in the experimental treatments. Either worms in test sediments experienced enhanced growth or the laboratory controls were stressed in some way. The latter explanation is likely as total control biomass was lower than previously observed (Johns et al. 1990) in such tests. However, the trend observed at the German Bight transect (Table 2) was similar to most other tests (decreasing toxicity away from sources).

Sediment toxicity tests used as part of the Bremerhaven Workshop provided essential information. They permitted smaller-scale spatial determinations than were possible for the majority of studies, which concentrated on dab *Limanda limanda*; for these other studies, for example, German Bight Stns 1 to 3 were too close to separate as were Stns 4 & 5. Similarly, at the drilling site, only the general area and a station 15 km away could be assessed using dab. Further, sediment toxicity tests were site specific and could be directly related to synoptic chemical measurements and coincident benthic community structure measures (Chapman 1992).

In the German Bight, sediment bioassays showed the same gradient as the chemistry. Use of chemical data alone would not have shown where toxicity occurred. Similarly, benthic community structure was not informative, as associations related to sediment physical differences and potential toxicant-related effects were not separable. Sediment toxicity tests conducted in the laboratory provide 'worst case' data. In the case of the drilling site, there was no obvious chemical gradient, and winter storms provided an overlying layer of clean sediment which had obscured any obvious effects on the meiofauna. The macrofaunal data suggested some changes close to the platform, but without supporting bioassay data these changes would have been suspect.

As noted by other investigators, sediment toxicity tests provide similar information for sites which are highly toxic (none of which were tested in this study), and different information for sites which are of low to moderate toxicity (such as those tested in this study). For instance, Becker et al. (1990) found that amphipod, oyster larvae and Microtox (elutriate rather than the direct exposure method used in this study) sediment toxicity tests could identify severely degraded communities although they had different associations with moderately impacted communities. Thus, determination of effects in intermediate areas requires 'burden of evidence' and 'best professional judgement' as recommended by Chapman (1991). This approach has been followed in this study and should be followed in future studies.

RECOMMENDATIONS

The results of the overall toxicity testing, considered above, indicate that sediments tested were not highly toxic and were, in fact, moderately to non-toxic. Thus, techniques were not tested across as wide a range of toxicity as is desirable (Mearns et al. 1986). However, based on the results of this study, the 10 d infaunal amphipod mortality and 48 h oyster larvae abnormality sediment toxicity test techniques are recommended for use in Europe.

Both tests should be conducted using standardised techniques, in particular preparation of elutriate for the oyster larval test, since this bioassay has been chosen for use in the North Sea Task Force Master Monitoring Plan. The Microtox direct exposure method shows promise, but remains to be fully developed (Dombroski et al. 1991, Munkittrick et al. 1991).

Other tests used in this study could be useful in future but require, at a minimum, further development and testing before they can be recommended for assessment and regulatory purposes in Europe. Some tests such as the clam reburial and oyster larvae metamorphosis, which can be done rapidly and relatively inexpensively on shipboard, could be useful for initial screening purposes if a range of responses can be demonstrated. However, bivalve embryos are more sensitive than older life stages (Ringwood 1991). The recommended toxicity test techniques could also benefit from further testing before they are used for regulatory purposes in European waters.

The above recommended tests have, as end-points, acute, primarily lethal responses. Chapman (1991) has recommended that organism-level responses to pollution include measures of survival, reproduction and growth. The one test used in this study which measured growth (*Neanthes arenaceodentata*) provided useful but unconvincing information; no tests used in this study measured reproduction. Swartz and co-workers (unpubl. data) have found that survival alone can give a greater than 70 % prediction of

population response, hence the immediate use of acute lethal tests provides a reasonable (but not complete) level of environmental protection. Accordingly, we recommend that future development of sediment toxicity tests include third generation tests as defined herein, namely those developed specifically for sediments which measure growth and reproduction as well as survival, and which preferentially do so over a full life-cycle. Alternatively, testing could use that portion of the life-cycle which is most appropriately sensitive following initial research to determine relative sensitivity of different life-cycle stages.

The *Neanthes arenaceodentata* growth test and/or other similar polychaete tests merit further research for use in Europe for several reasons. First, although control responses were questionable, worm growth along the German Bight field gradient was suggestive of a dose-response. Second, this test is a third generation sediment toxicity test, for which information on potential reproductive success may be available in future (Dr T. Dillon, U.S. ACOE/WES, pers. comm.). Third, polychaetes are numerically dominant members of the benthic community (Knox 1977). They are essential to the flow of energy, the cycling of nutrients, and can significantly affect the physio-chemical characteristics of sediments (Kristensen & Blackburn 1987, Meadows et al. 1990). Finally, in the North Sea, polychaetes are significant and quantitatively important members of the near-shore soft-bottom benthic community (Kristensen 1988, Esselink & Zwarts 1989, Menard et al. 1989, Zwarts & Esselink 1989).

Regardless of which species are used in toxicity tests, investigators should try to account for non-pollutant effects (e.g. grain-size) by choosing one or more appropriate reference sediments, or by removing such effects by regression as advocated by DeWitt et al. (1988). Further, the probability of declaring a sediment significantly different from control or reference depends on the quality of the control or reference data (as well as the number of sediments tested in certain statistical comparisons). As noted by Mearns et al. (1986), despite statistical significance, there is an area of uncertainty when survival is less than 90 % (but greater than 75 % for *Rhepoxynius abronius*), where sediments may or may not be toxic. In addition, as noted by Barnthouse et al. (1989), extremely precise data can result in very small and biologically inconsequential changes being found statistically significant which is often, incorrectly, equated directly with biological significance. Accordingly, interpretation and use of sediment toxicity test data, derived under 'worst case' laboratory exposure conditions, requires use of both best professional judgement and a burden of evidence approach, as discussed further by Chapman (1992).

LITERATURE CITED

Ankley, G. T., Katko, A., Arthur, J. W. (1990). Identification of ammonia as an important sediment-associated toxicant in the Lower Fox River and Green Bay, Wisconsin. Environ. Toxicol. Chem. 9: 313–322

ASTM (1986). Standard practice for conducting static acute toxicity tests with larvae of four species of bivalve molluscs. In: Annual book of ASTM standards, water and environmental technology, Vol. 11.04. American Society for Testing and Materials, Philadelphia, p. 368–384

Barnthouse, L. W., Suter, G. W. II, Rosen, A. E. (1989). Inferring population-level significance from individual-level effects: an extrapolation from fisheries science to ecotoxicology. In: Suter, G. W. II, Lewis, M. A. (eds.) Aquatic toxicology and environmental fate: 11th vol. ASTM STP 1007. American Society for Testing and Materials, Philadelphia, p. 289–300

Becker, D. S., Bilyard, G. R., Ginn, T C. (1990). Comparisons between sediment bioassays and alterations of benthic macroinvertebrate assemblages at a marine Superfund site: Commencement Bay, Washington. Environ. Toxicol. Chem. 9: 669–685

Bousfield, E. L. (1973). Shallow-water gammaridean Amphipoda of New England. Cornell University Press, New York

Butler, R., Chapman, P. M., van den Hurk, P., Roddie, B., Thain, J. E. (1992). A comparison of North American and West European oyster embryo-larval toxicity tests on North Sea sediments. Mar. Ecol. Prog. Ser. 91. 245–251

Chapman, P. M. (1986). Sediment bioassay tests provide toxicity data necessary for assessment and regulation. In: Geen, G. H., Woodward, K. L. (eds.) Proc. 11th Annual Aquatic Toxicity Workshop. Can. Tech. Rept. Fish. Aquat. Sci. 1480, p. 178–197

Chapman, P. M. (1988). Marine sediment toxicity tests. In: Lichtenberg, J. J., Winter, F A., Weber, C. I., Fradkin, L. (eds.) Chemical and biological characterization of sludges, sediments, dredge spoils, and drilling muds. ASTM STP 976. American Society for Testing and Materials, Philadelphia, p. 391–402

Chapman, P. M. (1991). Environmental quality criteria: what type should we be developing? Environ. Sci. Technol. 25: 1353–1359

Chapman, P. M. (1992). Pollution status of North Sea sediments – an international integrative study. Mar. Ecol. Prog. Ser. 91: 313–322

Chapman, P. M., Becker, S. (1986). Recommended protocols for conducting laboratory bioassays of Puget Sound sediments. U.S. Environmental Protection Agency, Puget Sound Estuary Program, Seattle

Chapman, P. M., Morgan, J. D. (1983). Sediment bioassays

with oyster larvae. Bull. environ. Contam. Toxicol. 34: 438–444

Chapman, P. M., Long, E. R., Swartz, R. C., DeWitt, T. H., Pastorok, R. (1991). Sediment toxicity tests, sediment chemistry and benthic ecology do provide new insights into the significance and management of contaminated sediments – a reply to Robert Spies. Environ. Toxicol. Chem. 10: 1–4

Cofino, W. P., Smedes, F., de Jong, S. A., Abarnou, A., Boon, J. P., Oostingh, I., Davies, I. M., Klungsøyr, J., Wilhelmsen, S., Law, R. J., Whinnett, J. A., Schmidt, D., Wilson S. (1992). The chemistry programme. Mar. Ecol. Prog. Ser. 91: 47–56

Daan, R., van het Groenewoud, H., de Jong, S. A., Mulder, M. (1992). Physico-chemical and biological features of a drilling site in the North Sea, 1 year after discharges of oil-contaminated drill cuttings. Mar. Ecol. Prog. Ser. 91: 37–45

DeWitt, T. H., Ditsworth, G. R., Swartz, R. C. (1988). Effects of natural sediment features on the phoxocephalid amphipod, Rhepoxynius abronius: implications for sediment toxicity bioassays. Mar. environ. Res. 25: 99–124

DiToro, D. M, Mahony, J. D., Hansen, D. J., Scott, K. J., Hicks, M. B., Mayr, S. M., Redmond, M. S. (1990). Toxicity of cadmium in sediment: the role of acid volatile sulfide. Environ. Toxicol. Chem. 9: 1489–1504

Dombroski, E. C., Gaudet, I. D., Qureshi, A. A. (1991). Review of the Microtox methodology for toxicity assessment of solid matrices. In: Chapman, P. M., Bishay, F. S., Power, E. A., Hall, K., Harding, L., McLeay, D., Nassichuk, M., Knapp, W. (eds.) Proc. 17th Annual Aquatic Toxicity Workshop, Vancouver, B.C., November 5–7, 1990. Can. Tech. Rept. Fish. Aquat. Sci. 1774, p. 478–494

Esselink, P., Zwarts, L. (1989). Seasonal trend in burrow depth and tidal variation in feeding activity of Nereis diversicolor. Mar Ecol. Prog. Ser. 56: 343–2254

EVS Consultants (1991). Phase IV studies of 10-day tests for sediment toxicity using marine or estuarine infaunal amphipods. Report prepared for Environment Canada, Marine Environment Division, Hull, Quebec

Gee, J. M., Austen, M., De Smet, G., Ferraro, T., McEvoy, A., Moore, S., Van Gausbeki, D., Vincx, M., Warwick, R. M. (1992). Soft sediment meiofauna community responses to environmental pollution gradients in the German Bight and at a drilling site off the Dutch coast. Mar. Ecol. Prog. Ser. 91. 289–302

Johns, D. M., Ginn, T. C., Reish, D. J (1990). Protocol for juvenile Neanthes sediment bioassay. U.S. Environmental Protection Agency, Puget Sound Estuary Program, Seattle. EPA 910/9–90–011

Knox, G. A. (1977). The role of polychaetes in benthic soft-bottom communities. In: Reish, D. J., Fauchald, K. (eds.) Essays on polychaetous annelids. Allan Hancock Foundation, Los Angeles, p. 547–604

Kristensen, E. (1988). Factors influencing the distribution of nereid polychaetes in Danish coastal waters. Ophelia 29: 127–140

Kristensen, E., Blackburn, T H. (1987). The fate of organic carbon and nitrogen on experimental marine sediment systems: influence of bioturbation and anoxia. J. mar. Res. 45: 231–257

Kröncke, l., Duineveld, G. C. A., Raak, S., Rachor, E., Daan, R. (1992). Effects of a former discharge of drill cuttings on the macrofauna community. Mar. Ecol. Prog. Ser. 91: 277–287

Kröncke, I., Rachor, E. (1992). Macrofauna investigations along a transect from the inner German Bight towards the Dogger Bank. Mar. Ecol. Prog. Ser. 91: 269–276

Long, E. R., Buchman, M. F., Bay, S. M., Breteler, D. J., Chapman, P. M., Hose, J. E., Lissner, A. L., Scott, J., Wolfe, D. A. (1990). Comparative evaluation of five toxicity tests with sediment from San Francisco Bay and Tomales Bay, California. Environ. Toxicol. Chem. 9: 1193–1214

Meadows, P. S. (1964). Substrate selection by Corophium species – the particle size of substrates. J. Anim. Ecol. 33: 387–394

Meadows, P. S., Tait, J., Hussain, S. A. (1990). Effects of estuarine infauna on sediment stability and particle sedimentation. Hydrobiologia 190: 263–266

Mearns, A., Swartz, R. C., Cummins, J., Dinnell, P., Chapman, P. M. (1986). Inter-laboratory calibration of a sediment toxicity test using the amphipod Rhepoxynius abronius. Mar. environ. Res. 18: 13–37

Menard, F., Gentil, F., Dauvin, J.-C. (1989). Population dynamics and secondary production of Owenia fusiformis Delle Chiaje (Polychaeta) from the Bay of Seine (eastern English Channel). J. exp. mar. Biol. Ecol. 133: 151–167

Munkittrick, K. R., Power, E. A., Sergy, G. A. (1991). The relative sensitivity of Microtox, daphnid, rainbow trout, and fathead minnow acute lethality tests. Environ. Toxicol. Water Qual. 6: 35–62

Phelps, H. L. (1989). Clam burrowing bioassay for estuarine sediment. Bull. environ. Contam. Toxicol. 43: 838–845

Phelps, H. L. (1990). Development of an estuarine sediment burrowing bioassay for shipboard use. Bull. environ. Contam. Toxicol. 45: 722–728

Phelps, H. L., Warner, K. A. (1990). Estuarine sediment bioassay with oyster pediveliger larvae (Crassostrea gigas). Bull. environ. Contam. Toxicol. 44: 197–204

Reish, D. J. (1984). Marine ecotoxicological tests with polychaetous annelids. In: Persoone, A., Jaspens, E., Claus, C. (eds.) Ecotoxicological testing for the marine environment, Vol. 1 State University of Ghent, Ghent, p. 427–454

Ringwood, A. H. (1991). Short-term accumulation of cadmium by embryos, larvae, and adults of an Hawaiian bivalve, Isognomon californicum. J. exp. mar. Biol. Ecol. 149: 55–66

Stebbing, A. R. D., Dethlefsen, V. (1992). Introduction to the Bremerhaven Workshop on Biological Effects of Contaminants. Mar. Ecol. Prog. Ser. 91: 1–8

Steel, R. G. D., Torrie, J. H. (1960). Principles and procedures of statistics. McGraw-Hill Book Co., New York

Swartz, R. C., DeBen, W. A., Phillips, J. K., Lamberson, J. O., Cole, F. A. (1985). Phoxocephalid amphipod bioassay for marine sediment toxicity. In: Cardwell, R. D., Purdy, R., Bahner, R. C. (eds.) Aquatic toxicology and hazard assessment: 7th symposium. ASTM STP 854, American Society for Testing and Materials, Philadelphia, p. 284–307

Swartz, R. C., Kemp, P. F., Schults, D. W., Ditsworth, G. R., Ozretich, R. O. (1989). Acute toxicity of sediment from Eagle Harbor, Washington, to the infaunal amphipod, Rhepoxynius abronius. Environ. Toxicol. Chem. 8: 215–222

Thain, J. E. (1991). Biological effects of contaminants: oyster (Crassostrea gigas) embryo bioassay. In: Techniques in marine environmental sciences No. 11, International Council for the Exploration of the Sea, Copenhagen

Thompson, T (1991). Summary of the sediment toxicity bioassay workshop, 17th Annual Aquatic Toxicity Workshop. In: Chapman, P. M., Bishay, F. S., Power, E. A., Hall, K., Harding, L., McLeay, D., Nassichuk, M., Knapp, W. (eds.) Proc. 17th Annual Aquatic Toxicity Workshop, Vancouver, B.C., November 5–7, 1990. Can. Tech. Rept. Fish. Aquat. Sci. 1774, p. 1051–1054

Tung, K-K., Scheibner, M. G., Walboum, C. C. (1991). The solid phase assay: a new Microtox test procedure. In: Chapman, P. M., Bishay, F. S., Power, E. A., Hall, K., Harding, L., McLeay, D., Nassichuk, M., Knapp, W. (eds.) Proc. 17th Annual Aquatic Toxicity Workshop, Vancouver, B.C., November 5–7, 1990. Can. Tech. Rept. Fish. Aquat. Sci. 1774, p. 495

van den Hurk, P., Chapman, P. M., Roddie, B., Swartz, R. C. (1992). A comparison of North American and West European infaunal amphipod species in a toxicity test on North Sea sediments. Mar. Ecol. Prog. Ser. 91: 237–243

van de Wassenberg, J. (1989). *Corophium* test as a bioassay for polluted marine sediments. Student report, State University of Utrecht (in Dutch)

Williams, L. G., Chapman, P. M., Ginn, T. C. (1986). A comparative evaluation of sediment toxicity using bacterial luminescence, oyster embryo, and amphipod sediment bioassays. Mar. environ. Res. 19: 225–249

Zar, J. H. (1984). Biostatistical analysis, 2nd edn. Prentice-Hall, Princeton, NJ

Zwarts, L., Esselink, P. (1989). Versatility of male curlews *Numenius arquatus* preying upon *Nereis diversicolor*: deploying contrasting capture modes dependent on prey availability. Mar. Ecol. Prog. Ser. 56: 255–269