



Review

The importance of both potency and mechanism in dose–response analysis: An example from exposure of Pacific herring (*Clupea pallasii*) embryos to low concentrations of weathered crude oil

Jerry M. Neff^a, David S. Page^{b,*}, Peter F. Landrum^c, Peter M. Chapman^d

^a Neff & Associates, LLC, 20 Templewood Drive, Duxbury, MA 02332-2946, USA

^b Bowdoin College, Chemistry Department, Brunswick, ME 04011, USA

^c 6829 Earhart Road, Ann Arbor, MI 48105-9712, USA

^d Golder Associates Ltd., 500–4260 Still Creek Drive, Burnaby, BC, Canada V5C 6C6

ARTICLE INFO

Keywords:

Polycyclic aromatic hydrocarbons

Toxicity

Herring embryos

Dose–response

Oil

ABSTRACT

This paper reanalyzes data from an earlier study that used effluents from oiled-gravel columns to assess the toxicity of aqueous fractions of weathered crude oil to Pacific herring embryos and larvae. This reanalysis has implications for future similar investigations, including the observance of two distinct dose–response curves for lethal and sublethal endpoints for different exposures in the same experiment, and the need to consider both potency and slope of dose–response curves for components of a toxicant mixture that shows potentially different toxicity mechanisms/causation. Contrary to conclusions of the original study, the aqueous concentration data cannot support the hypothesis that polycyclic aromatic hydrocarbons (PAHs) were the sole cause of toxicity and that oil toxicity increased with weathering. Confounding issues associated with the oiled gravel columns include changes in the concentration and composition of chemicals in exposure water, which interfere with the production of reliable and reproducible results relevant to the field.

© 2012 Elsevier Ltd. All rights reserved.

Contents

1. Introduction	8
2. Carls et al. (1999) methodology	8
3. Results and discussion	8
3.1. Condition of herring eggs	8
3.2. Chemistry and selection of dose metric	8
3.3. Exposure conditions	9
3.4. Toxicity responses	10
3.5. Confounding factors	11
3.6. Dose response	11
4. Conclusions	14
Acknowledgment	14
Appendix A. Supplementary material	14
References	14

* Corresponding author. Tel.: +1 207 729 4364.

E-mail addresses: neffjm@comcast.net (J.M. Neff), dpag@bowdoin.edu (D.S. Page), plandrumprof@charter.net (P.F. Landrum), Peter_Chapman@golder.com (P.M. Chapman).

1. Introduction

This paper examines the importance of considering both potency and mechanism of action of different chemicals in complex mixtures, such as crude oil, when analyzing dose–response relationships, particularly when comparing dose–response curves for biological response endpoints for exposures to mixtures with different compositions. Potency is defined as the probability of a dose having an adverse effect (Ryan, 1993). Changes in potency are most evident when the data in a multiple treatment study fail to follow a single or monotonic dose–response relationship, resulting in two or more discrete dose–response curves. Different mechanisms of toxicity can be implied when the slopes of the dose–response relationships for two exposures to complex hydrocarbon mixtures are different (Hayes, 2007). The absence of a monotonic dose–response relationship is indicative of the need to consider confounding factors including potentially unmeasured toxic compounds associated with the exposure methodology.

To examine this issue, we use, as a case study, experiments by Carls et al. (1999) that measured the effects of exposure of Pacific herring (*Clupea pallasii*) eggs (embryos) to aqueous extracts of crude oil that had undergone different degrees of weathering. This study provides a good example of the need to consider both potency and toxic mechanism when two distinct dose–response curves are obtained. In this study, Carls et al. (1999) concluded that low concentrations (0.4 µg/L) of dissolved total polycyclic aromatic hydrocarbons (TPAHs) from weathered crude oil were toxic to herring embryos and that weathering increased oil toxicity. These conclusions were based on a single set of un-replicated laboratory experiments. Although we have reviewed this work as well as a similar salmon study by Heintz et al. (1999) elsewhere (Page et al., 2012), we conducted a further review of this study because of the far reaching implications of the recommendation by Carls et al. (1999, 2002) that current water quality standards for PAH are not adequate to protect fish early life stages and the assertion that petroleum toxicity increases with weathering.

2. Carls et al. (1999) methodology

Carls et al. (1999) produced aqueous exposure media by pumping seawater up through vertical cylindrical columns containing gravel that had been coated with crude oil. This oil, which had been artificially weathered by heating overnight at 70 °C, was applied at four oil-on-gravel loading levels (trace, low, middle (mid), and high), plus control (no oil added). Prior to each experiment, gravid adult herring were collected in the field by Johnson et al. (1997) and artificially spawned in the laboratory. Adult herring for the LWO experiment were collected on April 11, 1995, at Cat Island near Ketchikan and adult herring for the MWO experiment were collected on May 13, 1995, at Seymour Canal, locations ~200 miles apart in southeast Alaska. One day post-fertilization herring embryos on glass slides were continuously exposed in exposure chambers to effluent water from the columns for 16 days. This regime was designated as the less weathered oil (LWO) experiment. At the end of the 16-day LWO experiment, water flow through the columns was stopped and surviving embryos were placed in clean seawater to continue development and for measurement of egg and larval survival and a group of sublethal responses. After 13 days, seawater flow was restarted in each column and a day later, a second batch of fertilized eggs was exposed to effluents from the same columns for 16 days; this experiment was designated as the more weathered oil (MWO) experiment.

PAH concentrations (41 individual PAH and alkyl-PAH congener groups) were measured in effluent water from all column oil loading levels and controls several times during the 16-day exposures.

PAH concentrations also were measured in embryos at days 4, 8, and 16 for all the MWO treatments as well as in day 1 and 2 embryos from the MWO-mid treatment. Embryos from only the LWO-high treatment, collected on days 4, 8, and 15 and after return to clean seawater on days 16, 17, 20, and 23, were analyzed for tissue PAH concentrations (Carls et al., 1997, 1999). The frequency of tissue analyses was unequal among treatments, sparse during the exposure phase of the experiments, and missing from the post-exposure phase of the study except for the LWO-high dose, making it difficult to interpret the accumulated dose associated with the toxic response.

3. Results and discussion

3.1. Condition of herring eggs

There were differences in the control mortalities of the eggs collected for the two experiments: ~5% for the LWO experiment and ~20% for the MWO experiment (Carls et al., 1999). These differences suggest that the health of the two batches of eggs was different for the LWO and MWO experiments. The high control mortality of eggs for the MWO experiment was just at the acceptable upper limit of 20% for chronic whole effluent toxicity studies (USEPA, 2002). The mean temperature of the MWO experiment was 1.1 °C higher than that for the LWO experiment (Carls et al., 1997) and mean salinity (32 psu) for both exposures was above the optimum range (12–17 psu) for incubation success of herring from southeast Alaska and British Columbia (Alderdice and Hourston, 1985). The differences in control mortality between the LWO and MWO studies suggest differences between the two studies related to both health of eggs and differences in the experimental conditions.

Differences in initial egg health were confirmed by the results of a concurrent study of reproductive success in herring by Johnson et al. (1997). This concurrent study used eggs collected at the same times and locations as the Carls et al. (1999) study. Johnson et al. (1997) report that the Seymour Canal embryos (same group as MWO experiment) hatched earlier because of the higher incubation temperature, had a lower percentage hatch and higher incidence of yolk sac edema than the Cat Island embryos (same group as LWO experiment), and noted that “environmental variation, such as extremes in temperature, salinity, dissolved oxygen, and ultraviolet radiation can also induce abnormalities in larval fish”.

Smith and Cameron (1979) reported a 10% incidence of gross abnormalities in Prince William Sound herring larvae 13 years prior to the Exxon Valdez oil spill, providing a baseline for the response parameters measured by Carls et al. (1999). The differences in the initial condition of the eggs in the two exposure experiments, the non-optimal incubation salinity, and the nature of the responses, which are not specific only to PAH toxicity but may result from a variety of stressors, may have influenced the experimental outcomes in an unpredictable manner and represent some of the confounding factors associated with this study.

3.2. Chemistry and selection of dose metric

Although Carls et al. (1999) quantified temporal concentration patterns of alkanes and PAH in water, tissue, and gravel samples, they assumed that all effects observed were caused by dissolved PAH in the column effluents. The only dose metric they used in their assessment was the initial aqueous concentrations of TPAH in column effluents. When performing a toxicity assessment, the selection of the dose metric is intended to relate directly to causality. Thus, by choosing TPAH as the dose metric, Carls et al. (1999)

implicitly assumed one of two likely scenarios: either that all PAH were contributing equally to mixture toxicity; or, that the TPAH contained the causative agent at concentrations proportional to the response. The latter assumption can be considered invalid for these experiments because there was not a constant relative concentration of the different PAH among the treatments, the result being that different treatments (aqueous doses) in each study were not simple dilution series of complex mixtures containing similar relative proportions of different oil PAH. In addition, the dynamics of the compound exposures were different for the various PAH, both within and among treatments, leading to a complex exposure regime (Landrum et al., 2013). Thus, the only reasonable rationale for selecting TPAH is the assumption of equal potency of all components of the complex petroleum mixture. However, there was no weighting of specific compounds in the mixture nor were groups of specific PAH evaluated as a sub-set of the data to support the subsequent hypothesis that high-molecular-weight PAH and alkyl-substituted PAH were the main contributors to effluent toxicity. In other words, the potency of specific PAH or groups of PAH was not established.

PAH are known to have a wide range of potencies and mechanisms of action, ranging from neutral narcosis (Di Toro et al., 2007; McGrath and Di Toro, 2009) to specific modes of toxic action (Billiard et al., 2008; Incardona et al., 2011). Thus, it is not surprising that a monotonic dose response was not found, particularly for endpoints related to the development of the embryos.

Carls et al. (1999) stated, but did not demonstrate, that all measured aqueous PAH were freely dissolved and none were associated with oil droplets, which leads to the assumption that all the individual PAH in exposure water were bioconcentrated independent of each other and other chemicals in the effluent. However, the analytical methods used by Carls et al. (1999) and in related studies, did not distinguish between freely dissolved and particulate oil (see Page et al., 2012). These assumptions are critical to the selection of TPAH as a dose metric and render the findings questionable because the effluents from the different oil-on-gravel loadings contained different initial concentrations and compositions of the measured alkanes and PAH that changed during both of the 16-day experiments. The presence of low solubility alkanes and high molecular weight alkyl PAH in the effluents from the oiled

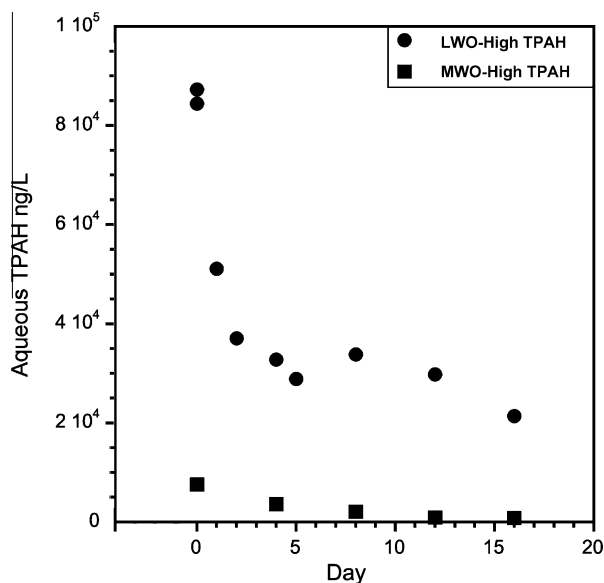


Fig. 1. Aqueous total polycyclic aromatic hydrocarbon (TPAH) concentrations (ng/L) in oiled-gravel-column effluents for the less weathered oil (LWO) and more weathered oil (MWO) high dose treatments.

Table 1

Measured total PAH (TPAH) concentrations ($\mu\text{g/L}$), relative aqueous concentrations (% analyte in TPAH), and measured concentrations ($\mu\text{g/L}$) of total ($\text{C}_1\text{--C}_4$) naphthalenes, phenanthrenes, and chrysenes at days 0 and 4 in effluents from the low, mid, and high loading columns for the less weathered oil (LWO) and more weathered oil (MWO) experiments^a. Shaded data (see text) are LWO-low and MWO-high data that have comparable initial TPAH values. and Data are from Carls et al. (1997, 1999) and EVOSTC (2009).

Treatment	Day	% Of TPAH		Concentration (µg/L)	
		LWO	MWO	LWO	MWO
Total polycyclic aromatic hydrocarbons (TPAHs)					
Low	0	100	100	9.1	0.43
	4	100	100	2.8	0.17
Mid	0	100	100	34	0.73
	4	100	100	20	0.73
High	0	100	100	86	7.6
	4	100	100	33	3.6
Total C ₁ –C ₄ -naphthalenes					
Low	0	61	12	5.6	0.05
	4	46	16	1.3	0.03
Mid	0	78	20	27	0.14
	4	72	20	14	0.14
High	0	77	59	66	4.5
	4	75	47	25	1.7
Total C ₁ –C ₄ -phenanthrenes					
Low	0	15	23	1.4	0.10
	4	25	29	0.7	0.05
Mid	0	3.9	31	1.3	0.23
	4	7.5	37	1.5	0.27
High	0	1.6	12	1.4	0.92
	4	5.2	18	1.7	0.66
Total C ₁ –C ₄ -chrysenes					
Low	0	0.08	2.1	0.01	0.01
	4	0.21	2.4	0.01	<0.01
Mid	0	0.03	0.76	0.01	0.01
	4	0.04	1.0	0.01	0.01
High	0	<0.01	0.21	0.01	0.02
	4	0.03	0.65	0.01	0.02
Total HMW PAH ^b					
Low	0	28	43	2.6	0.18
	4	41	55	1.2	0.09
Mid	0	12	48	4.0	0.36
	4	19	57	3.9	0.42
High	0	5.2	22	4.5	1.6
	4	14	30	4.7	1.1

^a Method detection limits generally were about 0.001–0.01 $\mu\text{g/L}$ per analyte or congener group.

^b Sum of quantified PAH (28 analytes) from phenanthrene through benzo(ghi)perylene.

gravel column studies (EVOSTC, 2009; Brannon et al., 2012; Page et al., 2012; Supplementary data) is indicative of the presence of a non-dissolved or micro droplet oil phase in the column effluents that probably contained all or most of the higher molecular weight PAH (Faksness et al., 2004; Redman et al., 2012). Therefore, the uptake and toxicity of PAH in the Carls et al. (1999) study likely cannot be attributed solely to a freely dissolved fraction of the oil PAH, and the likely presence of oil droplets represents an additional confounding factor that would affect the accumulated dose and that was not reported or discussed as part of the toxicology evaluation. Thus, total aqueous PAH, as measured, represented both freely dissolved and unknown amounts of PAH associated with oil droplets.

3.3. Exposure conditions

TPAH concentrations in exposure water declined rapidly and PAH composition changed continuously over the course of the 16-day exposures in all doses of LWO and MWO (Carls et al., 1997, 1999; EVOSTC, 2009; Supplementary data). The rapid decline of TPAH concentration in the LWO and MWO effluents during the

Table 2
Summary of day-0 aqueous polycyclic aromatic hydrocarbon (PAH) concentrations ($\mu\text{g/L}$) in the low, mid, and high doses of the less weathered oil (LWO) and the more weathered oil (MWO) effluents and associated statistically significant lethal and sublethal effects. Significant effects were determined by pairwise comparisons with controls for each treatment. Data are from Carls et al. (1997, 1999) and EVOSTC (2009).

Parameter	Less weathered oil (LWO)			More weathered oil (MWO)		
	Low	Mid	High	Low	Mid	High
Total PAH	9.1	34	86	0.43	0.73	7.6
Total HMW PAH ^a	2.6	4.0	4.5	0.18	0.36	1.6
C ₁ –C ₄ –naphthalenes	5.6	27	66	0.05	0.14	4.5
C ₁ –C ₃ –fluorenes	0.55	0.88	0.94	0.17	0.22	1.0
C ₁ –C ₄ –phenanthrenes	1.4	1.3	1.4	0.10	0.23	0.92
C ₁ –C ₃ –dibenzothiophenes	0.60	0.60	0.58	0.04	0.07	0.23
C ₁ –fluoranthenes/pyrenes	0.02	0.01	0.01	0.01	0.02	0.03
C ₁ –C ₄ –chrysenes	0.01	0.01	0.01	0.01	0.01	0.02
Total HMW alkyl PAH ^b	2.0	1.9	2.0	0.16	0.33	1.2
% HMW alkyl PAH	22	5.9	2.2	35	45	16
Significant embryo mortality?	N	Y	Y	N	N	Y
Significant larval mortality?	N	Y	Y	N	Y	Y
#Significant sublethal effects ^c	1	4	5	3	10	10

^a Sum of quantified PAH (28 analytes) from phenanthrene through benzo(ghi)perylene.

^b Sum of alkyl phenanthrenes, dibenzothiophenes, fluoranthenes/pyrenes, and chrysenes (rows 6–9).

^c 10 Sublethal effects were measured.

16-day exposures (Fig. 1) was largely the result of losses of lower molecular weight PAH, particularly naphthalene and alkyl-naphthalenes (Table 1). The relative concentrations of different individual PAH and PAH congener groups, as a percentage of TPAH concentration (%TPAH), changed in all effluent doses during the 16-day exposures, with percent alkyl-naphthalenes declining and percent alkyl-phenanthrenes, alkyl-dibenzothiophenes, and alkyl-chrysenes increasing in the low, mid, and high doses during the first 4 days of exposure (Table 1) and during the remainder of the two 16-day experiments. Thus, PAH exposure concentration declined and relative compositions were different for each dose during the course of the LWO and MWO experiments (EVOSTC, 2009).

Although the relative concentrations of alkyl-phenanthrenes and alkyl-chrysenes increased between days 0 and 4 in all doses in the LWO and MWO experiments, measured absolute concentrations decreased or increased only slightly (Table 1; Supplementary data). Relative and measured concentrations of alkyl-naphthalenes decreased in all doses except MWO-low and mid during the first 4 days of the 2 experiments. Thus, the declines in measured TPAH concentrations and changes in the relative PAH composition in the effluents at all doses of LWO and MWO were caused mainly by the more rapid loss of lower-molecular-weight naphthalene and alkyl-naphthalenes than the higher molecular weight (HMW) 3- and 4-ring parent and alkyl-PAH.

3.4. Toxicity responses

Carls et al. (1997, 1999) attributed the greater toxicity of the MWO effluent compared to the LWO effluent to the MWO's higher relative concentrations of HMW 3- and 4-ring parent and alkyl-PAH, in particular alkyl-phenanthrenes. However, it is the absolute concentration of a toxicant that determines toxic effects, not its relative concentration; again, the relative potency of the different HMW PAH should have been investigated. This is best illustrated by comparison of LWO and MWO doses with similar initial TPAH concentrations: LWO-low and MWO-high (bold face values in Table 1; see also Table 2). The LWO-low dose containing 9.1 $\mu\text{g/L}$ TPAH did not produce significant mortality in herring embryos (6.0%) and larvae (6.2%), whereas the MWO-high dose containing 7.6 $\mu\text{g/L}$ TPAH produced significant embryo and larval mortality (32.4% and 8.2% respectively). However, the non-toxic LWO-low effluent contained higher concentrations of TPAH, total HMW

PAH, total alkyl-naphthalenes, total alkyl-phenanthrenes and slightly lower concentrations of total alkyl-chrysenes than the toxic MWO-high effluent at both days 0 and 4 (Table 1). Total alkyl-chrysenes concentrations were comparable to analytical method detection limits in all effluents, including controls. Thus, the toxicity of the MWO-high effluent cannot be attributed to TPAH, total HMW PAH, alkyl-naphthalenes, alkyl-phenanthrenes, or alkyl-chrysenes.

In addition, the initial aqueous concentrations of TPAH, total HMW PAH, total HMW alkyl-PAH, total alkyl-naphthalenes, alkyl-phenanthrenes and alkyl-dibenzothiophenes in the MWO-low, mid, and -high doses that produced lethal and sublethal effects were lower than their concentrations in the LWO-low dose that was not lethal, but produced ~9% yolk sac edema in larvae (Table 2), comparable to the incidence of yolk sac edema in herring larvae from Seymour Canal (the source of eggs for the MWO experiment) (Johnson et al., 1997). Concentrations of alkyl-fluorenes, alkyl-fluoranthenes/pyrenes and alkyl-chrysenes were low in all doses although slightly higher in the MWO-high dose than in the

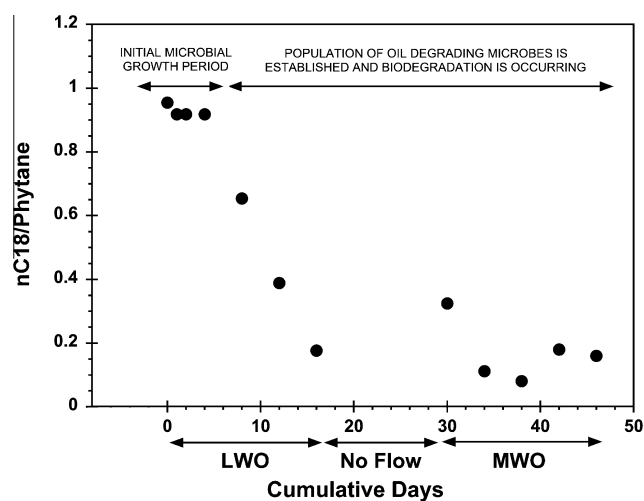


Fig. 2. Ratios of octadecane ($n\text{-C}_{18}$) to phytane concentration (an indicator of microbial biodegradation of oil) with time in the column effluent for the less weathered oil (LWO)-high and the more weathered oil (MWO)-high treatments. This ratio decreases as the more rapidly biodegraded $n\text{-C}_{18}$ is lost (NRC, 1985). Data from Carls et al. (1997, 1999) and EVOSTC (2009).

LWO-low dose. The initial TPAH concentrations in the different LWO and MWO doses (Tables 1 and 2) do not represent a simple dilution series or appropriate dose metric for oiled-gravel-column effluents, because the initial PAH composition of each dose was different from that of all other doses and the PAH compositions changed during the 16-day tests. Thus, the data of Carls et al. (1997, 1999) do not support the conclusion that the greater lethal and sublethal effects of the MWO effluents than the LWO effluents were caused by higher relative aqueous concentrations of HMW parent and alkylated PAH, because the measured concentrations of TPAH and different alkyl PAH congener groups in the toxic MWO doses were actually lower than in LWO doses that were not lethal and produced few sublethal effects.

3.5. Confounding factors

Because the oiled gravel columns were irrigated with unsterilized natural seawater and water flow was stopped for 13 days between the LWO and MWO studies, there was a strong potential for growth of hydrocarbon degrading microbes, resulting in biodegradation of petroleum hydrocarbon residues on the gravel (Wang et al., 1998) and microbial fouling of the eggs with production of microbial toxins as described by Grothe and Johnson (1996) and Hansen and Olafsen (1999). The ~35% decrease (from 21.4 to 7.6 µg/L) in the TPAH concentration in the column effluents between the day 16 LWO-high dose and the day 0 MWO-high dose (Carls et al., 1999), shown in Fig. 1, reflects a substantial loss of hydrocarbons during the 13 days between experiments when the water flow to the columns was stopped. The relative rates of depletion of readily biodegraded *n*-alkanes and of the less biodegradable branched alkanes, pristane and phytane, expressed as the *n*-C₁₇/pristane or *n*-C₁₈/phytane ratio, in the effluent from the oiled gravels are good indicators of microbial degradation of hydrocarbons (NRC, 1985; Kennicutt, 1988). These alkanes have extremely low aqueous solubilities, precluding depletion by dissolution from the oiled gravel columns. The *n*-octadecane (C₁₈)/phytane ratio is the more reliable indicator of oil biodegradation in marine environments because pristane is synthesized by some marine crustaceans and often is abundant in Arctic and sub-Arctic marine environments (Blumer et al., 1964). Pritchard et al. (1992) showed that the *n*-C₁₈/phytane ratio declined rapidly in weathered Exxon Valdez oil in the field in boulder/cobble sediments, even in the absence of added bioremediation fertilizer.

The *n*-C₁₈/phytane ratio in the LWO-high and MWO-high effluents declined rapidly during the respective experiments (Fig. 2) (EVOSTC, 2009; Supplementary data), indicating biodegradation of the more easily biodegraded *n*-C₁₈ (Wang et al., 1998). An oil-degrading microbial community apparently was established during the first 8 days of the LWO experiment, followed by extensive microbial degradation of oil on the columns during the remainder of the experiment (Fig. 2), as indicated by the rapid loss of *n*-C₁₈ between days 8 and 16. The ratio of *n*-C₁₈/phytane in the high dose LWO and MWO effluents decreased from 0.95 at day 0 to 0.08 at day 39, demonstrating extensive microbial degradation of oil on the gravel. Production of toxic microbial metabolites and degradation products of organic matter from all sources associated with the oiled gravel columns, as well as microbial fouling of the eggs, are additional unacknowledged confounding factors that could have contributed to effluent toxicity.

There also are reports of problems with microbial growth during the herring embryo experiments, recorded in the laboratory records from this study (Dahlberg, 1998). For example, the Carls Herring Study notebook p. 28, 6/5/95 notes: "Some jars are showing murky/milky/cloudy water. Filtrate stained for bacteria showed gram negative, chain forming bacteria—rods & cocci." Microbial activity, documented in some of the embryo incubation jars in the MWO

experiment, could have contributed to lethal and sublethal effects either directly or through the generation of toxic degradation products (see also Page et al., 2012). Middaugh et al. (1998, 2002) reported that microbial degradation of Alaskan North Slope crude oil produced toxic products, particularly in a polar subfraction of the water accommodated fraction (WAF), that were not present in the un-biodegraded WAF. The biodegraded crude oil produced developmental defects in inland silversides embryos similar to those reported by Carls et al. (1999) in herring embryos. The likely formation of toxic microbial metabolites and hydrocarbon degradation products during the two experiments contributes to the list of confounding factors in the Carls et al. (1999) study.

3.6. Dose response

Carls et al. (1999) established two aqueous dose–response curves for the LWO and MWO experiments, respectively, as shown in Carls et al. (1999) Figs. 4 and 5 for eight different lethal and sublethal responses. The occurrence of two dose–response curves based on the same dose metric invalidates a single cause-and-effect relationship based on that metric alone. This also makes it impossible to use this dose metric to predict responses under other exposures with this dose metric. In each of those figures, there are exposure doses in the LWO experiment which produced no effect for the same or greater aqueous TPAH exposure concentrations in the MWO experiment that produced effects.

Fig. 3A and B reproduces Fig. 4 of Carls et al. (1999) that shows the relationships between initial aqueous TPAH concentrations and mean percent mortality for herring embryos (A) and larvae at hatch (B). The PAH concentration/embryo mortality curves for initial TPAH (Fig. 3A) and for initial HMW PAH concentration/embryo mortality (Fig. 3D) show that, consistent with Table 1, embryo mortality for MWO treatments was observed at lower TPAH exposure concentrations (Fig. 3A) and lower HMW PAH concentrations (Fig. 3D) than for LWO treatments showing no embryo mortality, suggesting the lack of a clear causal link between TPAH concentration and the MWO effects observed.

Although Carls et al. (1999) present what appear to be two dose–response relationships for embryo and larval mortality (Fig. 3A and B), when one examines the mortality data for eggs, corrected for control mortality, there may only be a single dose response relationship for this endpoint. This might be expected as PAH are approximately equipotent (micromolar basis) for narcosis (Di Toro et al., 2007), which is often the mechanism for mortality. To examine this possibility, the data extracted from Carls et al. (1999) Fig. 4 were treated as data belonging to a single dose–response. For analysis, the data for MWO embryo mortality were corrected for control mortality using Schneider–Orelli's formula (Zeng et al., 2009), as recommended by the World Health Organization (WHO, 1998), because of the large difference in the control response between the LWO and MWO exposures. This correction was not required for the larval mortality because the control mortality was low and essentially equal in the two experiments. When corrected for the difference in control embryo mortality, the data in Fig. 3A appear to follow a single exposure concentration/response relationship (Fig. 3C). However, it is equally possible to retain the original two dose response curves, suggesting that differences in the factors controlling the mortality are likely from contributions from the confounding factors described above. Thus, the biological significance at low doses remains in question, because the LWO-low effluent at 9.1 µg/L TPAH did not produce egg mortality, whereas the MWO-high effluent caused approximately 17% mortality at 7.6 µg/L TPAH, after correction for control mortality (Fig. 3C). The confounding factors discussed above showing differences in the health of the eggs used in the LWO and MWO experiments probably contributed to the difference in the response

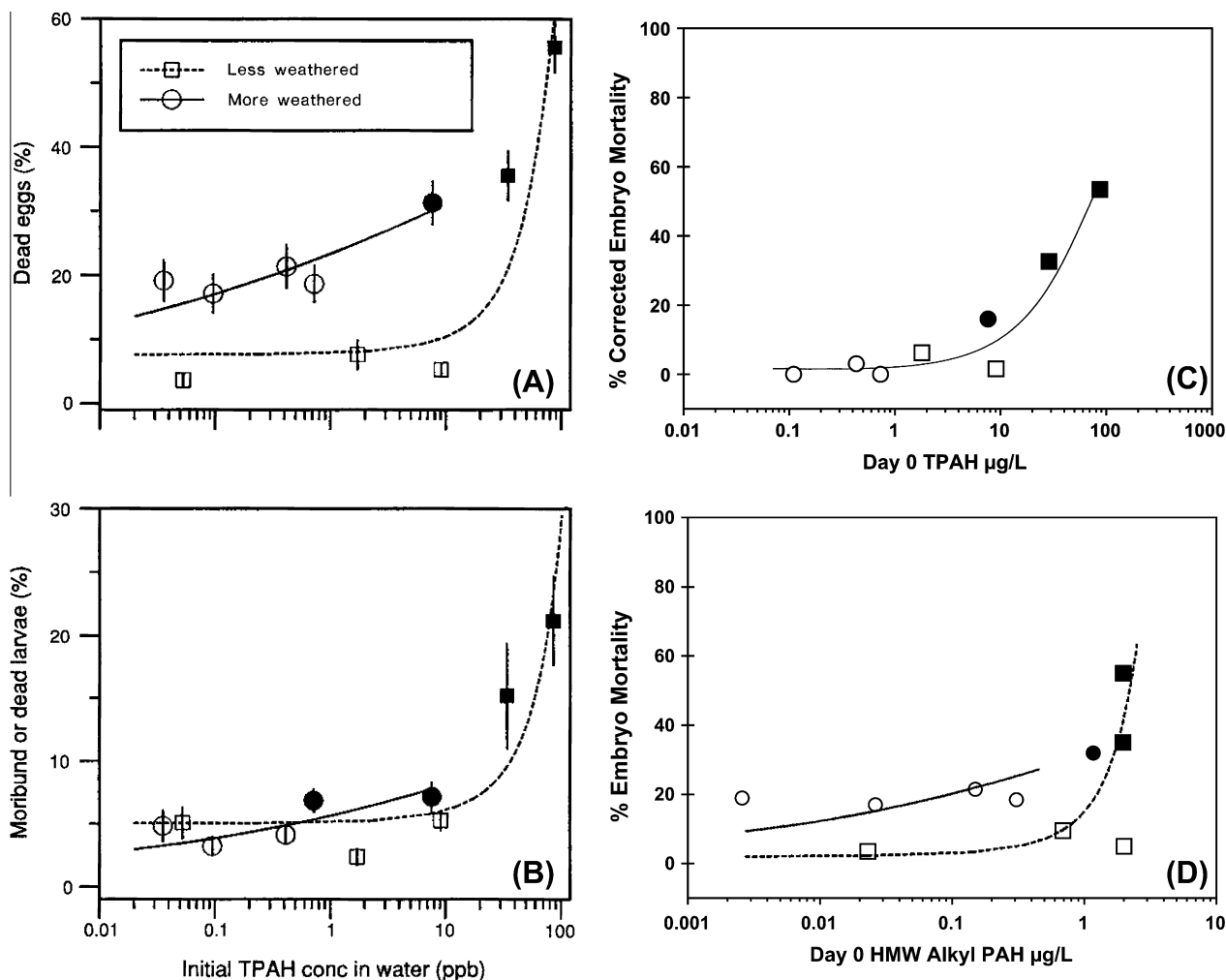


Fig. 3. Fig. 4 from Carls et al. (1999) showing embryo mortality and larval mortality at hatching as a function of initial aqueous TPAH concentration (ppb = $\mu\text{g/L}$) (A and B; with permission) for the less weathered oil (LWO) and more weathered oil (MWO) experiments. Embryo mortality data were extracted from (A), corrected, using Schneider-Orelli's formula (see text), for the large difference in control mortalities in the MWO and LWO experiments, and replotted (C). Initial concentrations of aqueous higher molecular weight (HMW) alkyl PAH, defined as the sum of alkyl-phenanthrenes, alkyl-dibenzothiophenes and alkyl-chrysenes, were plotted versus % embryo mortality (D). The lines in (D) reproduce the fits to the points in (A) and the line in (C) represents a logistic regression of all of the corrected data. Filled markers are significantly different from the respective control, with square markers representing LWO data and circles MWO data.

between the experiments. Other confounding factors likely also contributed.

Although it is possible to create a single dose response regression for the embryo toxicity data (Fig. 3C), this does not prove that aqueous TPAH (the chosen dose metric) are the only components of the column effluents contributing to the response, even though the observed response was approximately proportional to the initial TPAH concentration. Further, the PAH composition/concentration data for the nontoxic LWO-low and toxic MWO-high doses (Tables 1 and 2) also suggest that it is unlikely that a subfraction of PAH was substantially more potent than other subfractions for embryo mortality. This is confirmed by Fig. 3D, in which the HMW PAH, claimed by Carls et al. (1999) to be more potent than low MW PAH, show a similar overall concentration-response behavior to TPAH. What a single dose response does suggest is that the mechanism of action for mortality is likely consistent between the two experiments for mortality. Further, there may not be a strong potency gradient among the compounds because composition did not appear to play a role, which would suggest a relatively non-specific mechanism of toxic action (Di Toro et al., 2007). However, sorting out the contribution to the toxicity among the petroleum hydrocarbons and the degradation products is still required.

Thus, inference but not causality is established for the PAH subfraction of the petroleum mixtures (Landrum et al., 2012).

The concentration/response situation is completely different for the five most sensitive sublethal responses reported by Carls et al. (1999) in their Fig. 5. In that figure, there are some endpoints that show strong differences in potency as represented by the position of the LWO and MWO dose-response curves, and some that show both differences in potency and mechanism as represented by different slopes as well as different positions of the dose response curves. In all cases, the control responses are low; therefore, correction for control response would not have resulted in a single dose response curve unlike our finding for the embryo mortality above. Where the curves appear to be parallel and the MWO is shifted to the lower TPAH concentrations (e.g., pericardial edema, spinal defects, and effective swimmers), the simple presence of two dose-response curves demonstrates that the selection of TPAH as a dose metric is not adequate to describe the response. The driving force for such shifts in the dose-response can come from shifts in bioavailability, organism sensitivity, changes in mixture composition, and/or the presence of unknown toxicants acting by the same mechanism as suggested from the confounding factors outlined above. Unfortunately, there is inadequate information in

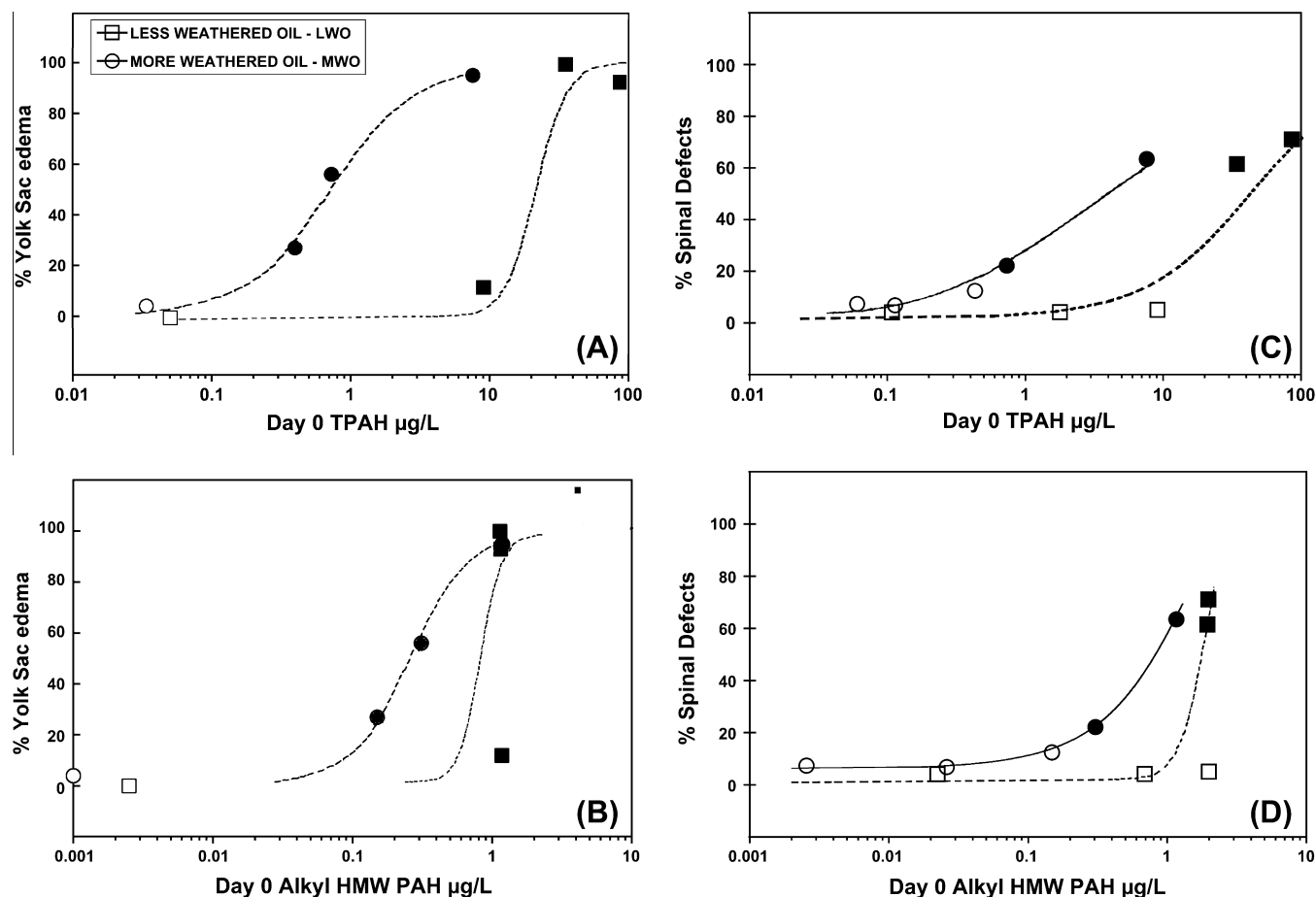


Fig. 4. Sublethal response data extracted from Fig. 5 of Carls et al. (1999), replotted as initial aqueous TPAH concentrations (EVOSTC, 2009) versus % yolk-sac edema (A) and % spinal defects in larvae (C) and as higher molecular weight (HMW) alkyl PAH concentrations, defined as the sum of alkyl-phenanthrenes, alkyl-dibenzothiophenes and alkyl-chrysenes concentrations ($\mu\text{g/L}$), versus % yolk-sac edema (B) and % spinal defects in larvae (D). The dotted lines (A–C) reproduce the fits to the points for the respective treatments given in Carls et al. (1999) Fig. 5 and are visually placed for D. Filled markers are significantly different from the respective control.

Carls et al. (1997, 1999), EVOSTC (2009); Dahlberg (1998) to sort out which are the primary factors contributing to these shifts.

Fig. 4 presents concentration–response data for 2 sublethal endpoints extracted from Fig. 5 of Carls et al. (1999) for both aqueous TPAH and for HMW alkyl-PAH exposures, which Carls et al. (1999) stated were responsible for the toxicity they observed. Fig. 4A and B shows the TPAH and HMW alkyl-PAH concentrations versus % larval yolk sac edema, a sublethal endpoint, a sublethal endpoint to be specific for exposure to PAH. However, yolk sac edema can originate from a variety of causes and is better considered a general indicator of stress (Page et al., 2012). Fig. 4C and D shows TPAH and HMW alkyl-PAH concentrations versus % spinal defects in hatched larvae as a sublethal endpoint, a general indicator of stress. Irrespective of the cause of the sublethal effects, the most important issue is the presence of two separate concentration–response curves for both sublethal responses, shown in Fig. 4 by the dotted lines traced from the fits to the points for the respective treatments from Carls et al. (1999).

When two different dose–response curves occur showing both a shift in potency and slope, the sublethal effect is almost certainly not due to a single causative factor. In this case, if toxicity had been due to TPAH alone (Fig. 4A and C) or to HMW alkyl-PAH (Fig. 4B and D), a consistent mechanism would have been expected, resulting in a single dose–response curve. Thus, the difference in the slopes of the dose–response relationships for the MWO and LWO exposures suggests different toxicity mechanisms for the same response. Changes in potency generally occur from different modify-

ing factors, as suggested above, whereas changes in slope (toxic mechanism) are generally thought to result from the presence of different toxicants acting by different mechanisms of action. Quantitative data on such modifying factors that could have contributed to changes in slope, such as the potential of microbial action either directly or through formation of metabolites as a potential cause were not available from this study to definitively address the source of the shift in the mechanism of action.

Thus, for sublethal endpoints, a convincing monotonic dose–response relationship was not established linking aqueous TPAH or HMW alkyl-PAH concentrations with observed toxicity. Reduced jaw, % effective swimmers, and pericardial edema, sublethal responses that were also reported by Carls et al. (1999) for all treatments, also show two dose–response relationships as occurred with larval yolk sac edema and spinal defects (Fig. 4) and show LWO data points with no toxicity at higher TPAH and HMW alkyl-PAH concentrations than MWO points that show a toxic effect. Although PAH are likely contributors to the observed sublethal responses, causation has not been established. Other chemicals in the effluents probably contributed to lethal and sublethal responses, particularly in the MWO experiment.

It is likely that PAH and alkane biodegradation products and microbial metabolites contributed to the toxicity of the column effluents, particularly for the MWO effluents. For example, some oxygenated PAH (microbial degradation products of PAH) are as toxic or more toxic than the metabolized PAH to early life stages of fish and produce sublethal effects, including yolk sac edema

and spinal defects, similar to those associated with exposure to complex mixtures of PAH (Carney et al., 2008; Fallahtafai et al., 2012). These biodegradation products would not be detected in water and tissues by the analytical methods used by Carls et al. (1999). Therefore, aqueous TPAH concentration would not be an accurate dose metric for these experiments if such materials are contributing significantly to the observed responses.

An assessment based on tissue residues, assuming that all toxicants were measured, might have led to a better understanding of the relationship between exposure and effects. However, a comparison across all treatments could not be performed because tissue PAH concentration data were not collected from all doses in the LWO study. Fig. 3 of Carls et al. (1999) suggests that the toxicokinetics for PAH in the two studies were substantially different on a wet-tissue-weight basis. Unfortunately, the absence of tissue data for all but the LWO-high dose treatment precludes determining whether a monotonic dose–response could have been established on a tissue PAH-residue basis between the LWO and MWO experiments for all endpoints.

Although PAH are expected to be significant contributors to the toxicity observed in these experiments, the actual contribution of each PAH compound in conjunction with PAH metabolites and other potential additional stressors identified as additional potential confounding factors to the different lethal and sublethal endpoints measured remains to be defined. The oiled-gravel columns produced effluents containing different concentrations and compositions of TPAH and total alkanes, proportional to the initial loading of oil on the columns. However, the initial relative concentrations of different PAH were not the same for the different treatments in the LWO and MWO experiments and the compositions changed in different ways during the two 16-day experiments because of different rates of depletion of PAH in the oil-on-gravel by dissolution, dispersion, and biodegradation. Therefore, it is not possible to determine the contribution of different PAH, alkanes, microbial degradation products and microbial fouling that led to the different lethal and sublethal endpoints observed. In addition, it is likely that the oiled gravel columns produced a mixture of dissolved and non-dissolved PAH (Page et al., 2012; Redman et al., 2012), further complicating the definition of aqueous exposure concentrations. Neither potency nor causation were determined by Carls et al. (1999) nor can they be determined based on the available data from this study.

The issues of causality and confounding factors identified here for the Carls et al. (1999) study have also been described for a similar study of pink salmon embryos and larvae (Landrum et al., 2012; Page et al., 2012). Given the rapidly declining aqueous PAH concentrations and variable aqueous PAH compositions produced by oiled gravel columns, it is not possible to define an aqueous PAH concentration causally associated with an observed effect (Landrum et al., 2013). This is particularly true for embryo toxicity tests, where embryos undergo rapid biochemical and morphogenic changes at the same time the exposure concentrations are declining and composition is changing most rapidly. This raises the question of whether the use of oiled-gravel columns to generate hydrocarbon-contaminated exposure media for toxicity studies can yield reliable and reproducible results that can be extrapolated to the field.

4. Conclusions

Toxicity studies need to demonstrate clear and convincing monotonic dose–response relationships between suspected toxicants and observed biological effects. The presence of two or more concentration–response relationships in multiple treatment studies is a strong indication of the presence of multiple stressors

and/or mechanisms of toxicity. Possible mechanistic explanations of the observed response or relationship need to be explored, including potency of other components of the complex mixtures, potential stressors, and possible confounding factors. Carls et al. (1999) did not demonstrate this and failed to consider the contribution of confounding factors, such as the use of adult herring collected from different locations and at different times as egg sources, microbial fouling of the oiled gravel, and associated production of toxic hydrocarbon oxidation products and microbial toxins. Because causality was not established, particularly with respect to the confounding factors, it is not possible to conclude that oil toxicity to herring embryos increases with weathering such that TPAH concentrations in the MWO effluents as low as 0.4 µg/L are toxic to herring larvae, when higher concentrations of the same TPAH in the LWO experiments produce no toxic effect. It is highly likely that unmeasured chemicals along with the confounding factors in the MWO effluents contributed to the observed toxicity. Thus, Carls et al. (1999) did not demonstrate that current water quality standards for TPAH are not adequately protective to fish early life stages. However, their study provides an excellent case study to illustrate the importance of both potency and mechanism in dose–response analysis. It also points out that the use of oiled-gravel columns to produce exposure media creates complex, rapidly changing mixtures of potential toxicants and has associated confounding issues that interfere with the production of reliable and reproducible results that can be extrapolated to the field.

Acknowledgment

Support for this work was provided by Exxon Mobil Corporation, Houston, TX; however, the conclusions are solely those of the authors and do not necessarily represent those of Exxon Mobil. We thank an anonymous referee and the journal editor for useful review comments.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.marpolbul.2012.12.014>.

References

- Alderdice, D.F., Hourston, A.S., 1985. Factors influencing development and survival of Pacific herring (*Clupea pallasii*) eggs and larvae to beginning of exogenous feeding. Can. J. Fish. Aquat. Sci. 42, 56–68.
- Billiard, S.M., Meyer, J.N., Wassenberg, D.M., Hodson, P.V., DiGiulio, R.T., 2008. Nonadditive effects of PAHs on early vertebrate development: mechanisms and implications for risk assessment. Toxicol. Sci. 105, 5–23.
- Blumer, M., Mullin, M.M., Thomas, D.W., 1964. Pristane in the marine environment. Helgol. Mar. Res. 10, 187–201.
- Brannon, E.L., Collins, K., Cronin, M.A., Moulton, L.L., Maki, A.W., Parker, K.R., 2012. Review of the Exxon Valdez oil spill effects on pink salmon in Prince William Sound, Alaska. Rev. Fish. Sci. 20, 20–60.
- Carls, M.G., Fremgen, D.M., Hose, J.E., Johnson, S.W., Rice, S.D., 1997. Effects of incubating herring (*Clupea pallasii*) eggs in water contaminated with weathered crude oil, and Appendices 4.1–4.5, in: Carls, M.G., Johnson, S.W., Thomas, R.E., Rice, S.D. (Eds.), Health and Reproductive Implications of Exposure of Pacific Herring (*Clupea pallasii*) Adults and eggs to Weathered Crude Oil, and Reproductive Condition of Herring Stock in Prince William Sound Six Years after the Exxon Valdez Oil Spill. National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Auke Bay Laboratory, Juneau, AK, USA, Exxon Valdez Oil Spill Restoration Project 95074 Final Report, 94–144, pp. 222–269 (Chapter 4).
- Carls, M.G., Marty, G.D., Hose, J.E., 2002. Synthesis of the toxicological impacts of the Exxon Valdez oil spill on Pacific herring (*Clupea pallasii*) in Prince William Sound, Alaska, USA. Can. J. Fish. Aquat. Sci. 59, 153–172.
- Carls, M.G., Rice, S.D., Hose, J.E., 1999. Sensitivity of fish embryos to weathered crude oil: Part I. Low level exposure during incubation causes malformations, genetic damage, and mortality in larval Pacific herring (*Clupea pallasii*). Environ. Toxicol. Chem. 18, 481–493.

- Carney, M.W., Erwin, K., Hardman, R., Yuen, B., Volz, D.C., Hinton, D.E., 2008. Differential developmental toxicity of naphthenic acid isomers in medaka (*Oryzias latipes*) embryos. *Mar. Pollut. Bull.* 57, 255–266.
- Dahlberg, M.L., 1998. Materials Provided Pertaining to National Oceanic and Atmospheric Administration Auke Bay Laboratory 1991–1995 Salmon and Herring Toxicological Studies. Freedom of Information Act Request #1998-0327, July 29, 1998 and December 3, 1999. National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Juneau, AK, USA.
- Di Toro, D.M., McGrath, J.A., Stubblefield, W.A., 2007. Predicting the toxicity of neat and weathered crude oil: toxic potential and the toxicity of saturated mixtures. *Environ. Sci. Chem.* 26, 24–36.
- EVOSTC, 2009. Exxon Valdez Oil Spill Trustee Council Exxon Valdez Hydrocarbon Database. Anchorage, AK, USA, Project 090290. Chemistry data extracted from this database are given at <<http://www.valdezsciences.com>>.
- Fakness, L.G., Grini, P.G., Daling, P.S., 2004. Partitioning of semi-soluble organic compounds between the water phase and oil droplets in produced water. *Mar. Pollut. Bull.* 48, 731–742.
- Fallahtafi, S., Rantanen, T., Brown, R.S., Sniekus, V., Hodson, P.V., 2012. Toxicity of hydroxylated alkyl-phenanthrenes to the early life stages of Japanese medaka (*Oryzias latipes*). *Aquat. Toxicol.* 106–107, 56–64.
- Grothe, D.R., Johnson, D.E., 1996. Bacterial interference in whole-effluent toxicity tests. *Environ. Toxicol. Chem.* 15, 761–764.
- Hansen, G.H., Olafsen, J.A., 1999. Bacterial interactions in early life stages of marine cold water fish. *Microb. Ecol.* 38, 1–26.
- Hayes, A.W. (Ed.), 2007. Principles and Methods of Toxicology, fifth ed. Informal Healthcare, London, UK, 1135 pp.
- Heintz, R.A., Short, J.W., Rice, S.D., 1999. Sensitivity of fish embryos to weathered crude oil: Part II. Increased mortality of pink salmon (*Oncorhynchus gorbuscha*) embryos incubating downstream from weathered Exxon Valdez crude oil. *Environ. Toxicol. Chem.* 18, 494–503.
- Incardona, J.P., Linbo, T.L., Scholz, N.L., 2011. Cardiac toxicity of 5-ring polycyclic aromatic hydrocarbons is differentially dependent on the aryl hydrocarbon receptor 2 during zebrafish development. *Toxicol. Appl. Pharmacol.* 257, 242–249.
- Johnson, S.W., Carls, M.G., Stone, R.P., Broderick, C.C., Rice, S.D., 1997. Reproductive success of Pacific herring, *Clupea pallasii*, in Prince William Sound, Alaska, six years after the Exxon Valdez oil spill. *Fish. Bull.* 95, 748–761.
- Kennicutt II, M.C., 1988. The effects of biodegradation on crude oil bulk and molecular composition. *Oil Chem. Pollut.* 4, 89–112.
- Landrum, P.F., Chapman, P.M., Neff, J., Page, D.S., 2012. Evaluating the aquatic toxicity of complex organic chemical mixtures: lessons learned from polycyclic aromatic hydrocarbons and petroleum hydrocarbon case studies. *Integr. Environ. Assess. Manage.* 8, 217–230.
- Landrum, P.F., Chapman, P.M., Page, D.S., Neff, J., 2013. Influence of exposure and toxicokinetics on measures of aquatic toxicity for organic contaminants: a case study review. *Integr. Environ. Assess. Manage.* <http://dx.doi.org/10.1002/ieam.1388>.
- McGrath, J.A., Di Toro, D.M., 2009. Validation of the target lipid model for toxicity assessment of residual petroleum constituents: monocyclic and polycyclic aromatic hydrocarbons. *Environ. Toxicol. Chem.* 28, 1130–1148.
- Middaugh, D.P., Shelton, M.E., McKenney, C.L., Cherr, G., Chapman, P.J., Courtney, L.A., 1998. Preliminary observations on responses of embryonic and larval Pacific herring, *Clupea pallasii*, to neutral fraction biodegradation products of weathered Alaska North Slope oil. *Arch. Environ. Contam. Toxicol.* 34, 188–196.
- Middaugh, D.P., Chapman, P.J., Shelton, M.E., McKenney Jr, C.L., Courtney, L.A., 2002. Effects of fractions from biodegraded Alaska North Slope crude oil on embryonic inland silversides, *Menidia beryllina*. *Arch. Environ. Contam. Toxicol.* 42, 236–243.
- NRC (National Research Council), 1985. Oil in the Sea: Inputs, Fates, and Effects. National Academy Press, Washington, DC, USA.
- Page, D.S., Chapman, P.M., Landrum, P.F., Neff, J., Elston, R., 2012. A perspective on the toxicity of low concentrations of petroleum-derived polycyclic aromatic hydrocarbons to early life stages of herring and salmon. *Human Ecol. Risk Assess.* 18, 229–260.
- Pritchard, P.H., Mueller, J.G., Rogers, J.C., Kremer, F.V., Glaser, J.A., 1992. Oil spill bioremediation: experiences, lessons and results from the Exxon Valdez oil spill in Alaska. *Biodegradation* 3, 315–335.
- Redman, A., McGrath, J., Stubblefield, W., Maki, A., Di Toro, D., 2012. Quantifying the concentration of crude oil microdroplets in oil-water preparations. *Environ. Toxicol. Chem.* 31, 1814–1822.
- Ryan, L., 1993. Potency measures for developmental toxicity. *Environmetrics* 4, 507–518.
- Smith, R.L., Cameron, J.A., 1979. Effect of water soluble fraction of Prudhoe Bay crude oil on embryonic development of Pacific herring. *Trans. Am. Fish. Soc.* 108, 70–75.
- USEPA (US Environmental Protection Agency), 2002. Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms, fourth ed. EPA-821-R-02-013, Office of Water, Washington, DC, USA, 335 pp.
- Wang, Z., Fingas, M., Blenkinsipp, S., Sergy, G., Lardriault, M., Sigouin, L., Foght, J., Semple, K., Westlake, D.W.S., 1998. Comparison of oil composition changes due to biodegradation and physical weathering in different oils. *J. Chromatogr. A* 809, 89–107.
- WHO (World Health Organization), 1998. Test Procedures for Insecticide Resistance Monitoring in Malaria Vectors, Bio-efficacy and Persistence of Insecticides on Treated Surfaces. WHO/CDS/CPC/MAL/98.12, Geneva, Switzerland.
- Zeng, Z., Richardson, J., Verdusco, D., Mitchell, D.L., Patton, E.E., 2009. Zebrafish have a competent p53-dependent nucleotide excision repair pathway to resolve ultraviolet B-induced DNA damage in the skin. *Zebrafish* 6, 405–415.