Critical Review

Oil Toxicity Test Methods Must Be Improved

Peter V. Hodson,a,b,* Julie Adams,a and R. Stephen Browna,c

a School of Environmental Studies, Queen’s University, Kingston, Ontario, Canada
b Department of Biology, Queen’s University, Kingston, Ontario, Canada
c Department of Chemistry, Queen’s University, Kingston, Ontario, Canada

Abstract: A review of the literature on oil toxicity tests showed a high diversity of reported test methods that may affect the composition, stability, and toxicity of oil solutions. Concentrations of oil in test solutions are dynamic because hydrocarbons evaporate, partition to test containers, bioaccumulate, biodegrade, and photo-oxidize. As a result, the composition and toxicity of test solutions may vary widely and create significant obstacles to comparing toxicity among studies and to applying existing data to new risk assessments. Some differences in toxicity can be resolved if benchmarks are based on measured concentrations of hydrocarbons in test solutions, highlighting the key role of chemical analyses. However, analyses have often been too infrequent to characterize rapid and profound changes in oil concentrations and composition during tests. The lack of practical methods to discriminate particulate from dissolved oil may also contribute to underestimating toxicity. Overall, current test protocols create uncertainty in toxicity benchmarks, with a high risk of errors in measured toxicity. Standard oil toxicity tests conducted in parallel with tests under site-specific conditions would provide an understanding of how test methods and conditions affect measured oil toxicity. Development of standard test methods could be achieved by collaborations among university, industry, and government scientists to define methods acceptable to all 3 sectors.

INTRODUCTION

The present critical review examines factors that affect the outcome, interpretation, and validity of oil toxicity tests. Toxicity tests are essential components of ecological risk and impact assessments of spilled oil. However, results are highly variable as a result of the diversity of test designs and methods that affect measured toxicity. To understand the risks of oil spills to aquatic ecosystems, there is a need for standard methods for storing and handling oil, preparing and characterizing test solutions, exposing aquatic species, and reporting test conditions.

There are significant challenges in establishing unambiguous measures of the inherent toxicity of oil. These include complex mixtures of volatile, hydrophobic, and biodegradable hydrocarbons, many of which can be photo-oxidized. Oil concentrations in test solutions are inherently unstable and sensitive to experimental methods that affect the amount, bioavailability, and estimated toxicity of dissolved hydrocarbons. Improved protocols are needed to support reliable and repeatable estimates of toxicity and comparisons among oils and test conditions (Aurand and Coelho 2005; Redman and Parkerton 2015).

Another challenge is the diversity of ecosystem characteristics that affects the fate and behavior of hydrocarbons at spill sites, modifies the exposure and responses of aquatic biota, and determines site-specific risks. Ecological risk assessments demand realistic tests of toxicity under conditions typical of predicted or actual spill sites (Bejarano et al. 2014). However, diverse test conditions hinder the application of site-specific toxicity data to risk assessments for other sites.

The present review evolved from a literature survey of toxicity test methods for crude and refined oils (Adams et al. 2017) that identified a wide array of test methods that affect measured toxicity. The objectives of the present review are to identify which test methods affect the outcome of toxicity tests and therefore should be standardized, to identify research needs to support the development of better methods, and to promote method standardization by collaborations among multiple stakeholders.
TOXICITY TEST REQUIREMENTS

A toxicity test is a determination of the effect of a material on a group of selected organisms under defined conditions (Environment Canada 2007). Toxicity tests are controlled experiments in which groups of organisms are exposed to a gradient of concentrations of a test substance to define accurate and precise endpoints such as a median lethal concentration (LC50) or a median effect concentration (EC50). To achieve these goals, each test of oil toxicity should satisfy a series of rarely stated assumptions including: 1) all test variables are constant except a gradient of oil concentrations, 2) oil concentrations decrease in proportion to loadings or dilutions of test solutions, 3) each test concentration is known and constant throughout the test, 4) petroleum hydrocarbons are freely dissolved in test solutions, and 5) the constituents of oil remain in constant proportion to each other over time and among test concentrations.

Some assumptions were derived from Sprague (1969), whose recommendations for “best practice” were adopted in formal test guidelines (e.g., Environment Canada 2007). Redman and Parkerton (2015) provided additional principles for acute lethality tests with oil, including that toxicity is determined by the concentration and composition of dissolved hydrocarbons, which depend in turn on the composition of test oils and methods for preparing oil solutions.

The Organisation for Economic Co-operation and Development (OECD) (2018) supplied practical strategies to test “difficult substances and mixtures” and to systematically mitigate the effect of test conditions on the stability and toxicity of test solutions before conducting definitive tests. Meeting test assumptions by standardizing toxicity test methods will reduce uncertainty in measured toxicity, calculated benchmarks, and estimated ecological risks.

Recommendation

The OECD guidance document on aqueous-phase aquatic toxicity testing of difficult test chemicals should be applied to improve and standardize oil toxicity test methods.

HOW UNIFORM ARE OIL TOXICITY TEST METHODS?

Adams et al. (2017) reviewed 144 papers reporting on oil toxicity; the majority of the papers were published since 2010 and dealt with fish. The most notable findings were the wide array of test objectives, experimental designs (i.e., exposure regimes), methods for preparing and characterizing test solutions and exposing test organisms, and experimental details reported (see also Table 1 in Redman and Parkerton 2015). For example, 185 tests with oil applied 223 different methods. Diversity was often a product of “re-inventing the wheel” without reference to published methods, and a focus on site-specific test conditions without comparing site-specific toxicity with a benchmark of toxicity established under standard conditions.

Many papers cited the Chemical Response to Oil Spills: Ecological Effects Research Forum (CROSERF) protocols that were developed through a very effective consensus-based initiative of government and industry scientists (Aurand and Coelho 2005). Similar to the present review, the intent of CROSERF was to reduce interlaboratory variations among toxicity tests of mechanically dispersed oil (water-accommodated fractions [WAF]) and chemically dispersed oil (chemically enhanced water-accommodated fractions [CEWAF]). The recommendations of the present review echo many of the CROSERF themes, although some proposals for test methods may differ. The CROSERF guidelines emphasized mixing protocols, exposure regimes, and chemical analyses of test solutions to produce data relevant to open-water marine oil spills. Although the guidelines were often cited as the source of test methods, most studies deviated from the recommended methods without presenting a rationale for the deviations or appropriate controls to show how those deviations affected the composition and toxicity of test solutions. Thus variations among studies in reported oil toxicity may really represent variations in methods applied.

The wide array of test methods was reflected in a diversity of terminology. Anderson et al. (1974) defined widely quoted procedures to prepare a water-soluble fraction (WSF) of oil, a term implying that all hydrocarbons were dissolved, and that particulate oil was absent. Whereas this might be true if droplets were removed from test solutions, few studies verified that droplets were truly absent. The WAF designation, which may include both dissolved and particulate oil (Redman et al. 2017; Sandoval et al. 2017), is a broader and more suitable description of test solutions. Many other terms were used to describe oil solutions, and most were less well defined than WAF, WSF, or CEWAF (Table 1). Some designations were variations on CROSERF terms, with qualifiers to describe how solutions were prepared such as WAF generated by low energy (LEWAF), medium energy (MEWAF), or high energy (HEWAF) mixing. Other studies applied CROSERF terms to specific oils, introduced new methods (e.g., biologically enhanced WAF), referred to WSF when researchers had really prepared WAF, or introduced unique and obscure terms (e.g., oil-only media). Diverse terminology does not foster a ready sharing of test data among risk or impact assessments, no matter how well the experiments were conducted. In the present review, WAF describes physically dispersed oil and CEWAF identifies chemically dispersed oil.

TEST METHODS—STANDARD VERSUS SITE-SPECIFIC

The strongest argument for standard test methods is that benchmarks of oil toxicity may be in error to an unknown degree and in an undefined direction if the effects of methods on test outcomes are undetermined. For example, species sensitivity distributions (SSDs) are relied on for ecological risk assessments (Bejarano et al. 2014). They provide perspective on the relative sensitivity to oil among test species and statistically derived benchmarks such as the oil concentration affecting the most sensitive 5th percentile of species tested. The SSDs for single compounds derived from standard toxicity tests can be quite reliable. However, for complex mixtures such as oil, each data
point in the SSD includes an unknown variance associated with the: 1) oil tested (each a unique mixture of hydrocarbons), 2) degree of oil weathering, 3) unique patterns of exposure associated with experimental conditions, 4) decline of concentrations over time, 5) amount of particulate oil in test solutions, and 6) extent to which test solutions were characterized by chemical analyses. Depending on the size and direction of measurement errors, the rank order of species sensitivity and derived benchmarks may vary considerably.

Standard test methods provide a broader and more reliable database for risk and impact assessments. Nevertheless, some variations among tests are unavoidable because each crude oil has unique physical and chemical characteristics. Of the 75 different oils described in the papers reviewed by Adams et al. (2017) more than 50 were crude oils and the remainder were light and heavy fuel oils. Marine and freshwater toxicity tests require different water qualities and test species, as do tests relevant to tropical, temperate, or arctic ecosystems. Thus standard methods must be sufficiently flexible to accommodate conditions relevant to different climates and salinities.

There is also a critical need to understand how toxicity changes under the conditions of a spill site, such as the rapid dilution of oil after open water spills (Bejarano et al. 2014). However, no laboratory experiment is realistic given the number of physical, chemical, and ecological characteristics that are unique to each spill site, the inherent variability of each characteristic, and uncontrollable variables such as weather and sea state. As the number of test variables increases, the relevance of test data to other spill scenarios decreases, whereas the complexity of modeling and predicting toxicity may become unmanageable.

The value of both standard and site-specific approaches may be realized by combining them (Aurand and Coelho 2005). By measuring toxicity under standard conditions, perspective can be gained on the relative hazard of each type of oil. By repeating the test under site-specific conditions, the influence of each condition on measured toxicity can be evaluated systematically. Standard methods can also be flexible in their application without losing their integrity. For instance, the duration of acute lethality tests is usually 24 to 96 h, depending on the species tested. If shorter tests are needed to assess the effects of brief exposures (e.g., a 4-h LC50), mortality can be recorded at short intervals from start to 96 h (Sprague 1969). Similarly, latent or delayed effects can be assessed by transferring survivors to clean water for extended observation periods. Hence a single test can generate data for risk assessments of standard conditions, brief exposures, and latent effects post-exposure; nevertheless, extra treatments may be needed to assess latent effects of brief exposures.

**Recommendation**

Standard methods for oil toxicity tests should be developed and applied in tandem with tests under site-specific conditions to provide perspective on test conditions that enhance or reduce measured toxicity.

### EXPOSURE REGIMES

Exposure regimes include the planned or nominal range of oil concentrations, plus unplanned changes during the test. The most commonly reported regimes were static (no renewal of test solutions throughout the experiment, 36.5% of 167 studies),
semi static (e.g., daily renewal, 26.9%), and continuous-flow (constant addition of oil to flowing water, 26.3%; Adams et al. 2017). One CROSERF regime models open-water oil dispersion by adding a continuous flow of clean water to oil-contaminated water to reduce oil concentrations from maximum to less than analytical detection limits in a fixed interval (Aurand and Coelho 2005; Bejarano et al. 2014). However, this spike scenario is unsuitable for spills where oil does not dissipate rapidly, including confined ecosystems (rivers, lakes, and marshes) or the continuous discharge of oil from a point source (pipeline break or oil well blowout).

Constant concentrations of oil in test solutions cannot be sustained during static and semi-static exposure regimes, particularly when solutions contain oil droplets. After oil and water are mixed, oil droplets coalesce and rise to the surface, small compounds (<C12) evaporate, and hydrophobic compounds partition from water to test organisms, to the walls of test containers, and to residual oil droplets (Sandoval et al. 2017). For example, concentrations of polycyclic aromatic compounds (PAC) in static solutions of crude oil HEWAF declined exponentially by 90% in 24 h. For CEWAF the decrease was only 26%, reflecting the stabilizing effects of dispersants, and for WAF the reduction was only 6% (Sandoval et al. 2017). However, these experiments did not include test organisms that would likely increase loss rates. When test solutions are renewed daily, exposure regimes may represent a series of spikes, with sharp deviations from nominal concentrations (Figure 1). Under these conditions, it is unknown whether test organisms respond to initial (time 0), final, or average oil concentrations, or to some integral of time and concentration such as the “time-weighted mean exposure concentration” recommended by the Organisation for Economic Co-operation and Development (2018). Most importantly, when endpoints are calculated from nominal oil concentrations, or concentrations measured at time 0, toxicity can be underestimated by more than 10-fold, as shown with a single PAC (Kiparissis et al. 2003). Although static or semi-static trials are simple and convenient, they do not necessarily provide accurate benchmarks of oil toxicity.

Alternative exposure regimes include continuous-flow dosing systems. To illustrate, fine oil droplets injected into flowing water by high-pressure pumps and nozzles create high concentrations of dissolved hydrocarbons (Nordtug et al. 2011); residual droplets are removed by an oil–water separator before dilution for testing. This system provides constant oil concentrations for prolonged experiments but requires high volumes of water and oil. On a smaller scale, fresh or weathered oil was mixed with water by peristaltic pumps and in-line mixers (Croce and Stagg 1997), although interactions between oil and plastic peristaltic tubing may limit this approach to brief experiments.

Oil has also been added directly to exposure tanks in floating corrals with re-circulation pumps to enhance oil–water partitioning and a continuous input of hydrocarbons (Danion et al. 2011). However, surface oil weathered rapidly with a 75% loss of initial concentrations and a change in the hydrocarbon composition of test solutions.

Constant concentrations of oil in static and continuous flows of WAF can also be created by passive dosing. For instance, as water passes through columns of oil-coated substrates, hydrocarbons partition from oil films to water, and concentrations of hydrocarbons may reach (but not exceed) their solubility limits when the coating is fresh (Heintz et al. 1999; Martin et al. 2014). This system sustains measurable concentrations of oil in water for much longer than static or semi-static regimes. However, it models exposure not to fresh oil but to stranded, weathered oil. The composition and concentrations of test solutions change over days and weeks as the more water-soluble components are depleted (Carls et al. 1999; Martin et al. 2014) and when droplets are periodically released from oil films (J. Adams, unpublished data).

Passive dosing for static test solutions is equivalent to the partition-controlled delivery of a single PAC by equilibrium partitioning from polydimethylsiloxane (silicone) films (e.g., Kiparissis et al. 2003; Lin et al. 2015). For oil solutions, passive dosing systems include silicone tubing filled with oil (Redman et al. 2017) or silicone O-rings soaked in oil (Bera et al. 2018). Hydrocarbons partition from oil to silicone and from silicone to water when the silicone media are immersed in water with constant stirring (Redman et al. 2017). Equilibria are reached in 48 to 72 h, after which aqueous concentrations are sustained despite loss pathways such as bioaccumulation. Passive dosing appears technically simpler than continuous-flow technologies and requires less oil. Times to equilibria and the concentrations of oil transferred to water compared favorably with WAF solutions produced by stirring water and oil for up to 72 h (Bera et al. 2018). For all passive dosing systems, test solutions should be analyzed frequently to describe changes in composition as the more water-soluble and degradable components are depleted (Organisation for Economic Co-operation and Development 2018), and to detect the release of oil droplets (Redman et al. 2017).

Continuous-flow and passive dosing methods are not yet widely used. Research is needed to show their effectiveness under different test conditions (e.g., Organisation for Economic Co-operation and Development 2018), and to ensure that assumptions about toxicity tests are met. (See Toxicity Test Requirements section).
Recommendations

Continuous-flow and passive dosing exposure regimes should be adopted for standard oil toxicity tests and for all exposure systems; the stability of test solutions should be assessed before and during the tests.

Research needs

1) Identify the optimal conditions for the continuous addition of oil to water.
2) Develop statistical models for both static and semi-static exposure regimes that integrate oil concentrations in test solutions over time to generate stable metrics of exposure and toxicity.

PREPARING WAF AND CEWAF SOLUTIONS

Until alternative methods have been optimized and are widely accepted, static and semi-static protocols will continue to be applied. Therefore, it is important to understand how each aspect of oil mixing and dispersion for WAF and CEWAF affects the composition and concentration of oil in test solutions, particularly if methods vary among studies. Adams et al. (2017) found little consistency across the literature in all aspects of mixing methods, and details on mixing were often not reported at all (Supplemental Data, Table S1).

For example, in the CROSERF method each concentration of WAF or CEWAF was mixed independently in the volume required for the test (Aurand and Coelho 2005); gradients of concentration were prepared with gradients of oil to water ratio mixed by large-scale stirrers (Bejarano et al. 2014). However, gradients of oil concentration also included gradients of hydrocarbon composition; hence the ratio of low to high molecular weight hydrocarbons in the solution increased with the oil to water ratio. Barron and Ka’aihue (2003) proposed that small volumes of stock solution be prepared at a single oil to water ratio, followed by variable dilutions to create concentration gradients with a fixed hydrocarbon composition. This approach has been widely accepted and is recommended in the present review to simplify the preparation of test solutions and to avoid the confounding effects of concentration–dependent hydrocarbon composition.

Another well-known example is the dependence of the acute lethality of oil on the head space in closed mixing vessels. For fresh oil, toxicity is greatest when the head space is small because the low volatility compounds associated with acute effects are retained in closed vessels (Echols et al. 2016). Although some studies applied an appropriate head space, more than 90% failed to report head space at all (Table 1), creating questions about the accuracy of reported toxicity.

As reviewed in the Test Methods—Standard versus Site-Specific section, the amount of dissolved and particulate oil in test solutions of WAF, CEWAF, and HEWAF determines the rate of oil–water partitioning, the concentration of dissolved hydrocarbons, and the toxicity of test solutions (Nordtug et al. 2011; Pan et al. 2017). High energy WAF (HEWAF) is more toxic than WAF because higher mixing energies disperse droplets that are a source of dissolved hydrocarbons in test solutions; thus mixing energy is a critical determinant of toxicity (Sandoval et al. 2017). Although many acronyms imply high or low energy mixing (Table 1), there are no published methods to measure and replicate mixing energy in flasks. Indirect estimates are possible by particle image velocimetry and velocity field modeling (Zhao et al. 2016); however, most studies follow a rule of thumb (~ vortex depth during stirring) to standardize mixing energy among tests (Aurand and Coelho 2005). Given the variety of reported mixing conditions that could affect mixing energy (stirring speed, duration, oil-to-water ratio, mixing vessel shape, and settling time; see Supplemental Data, Table S1), it is likely that concentrations of particulate and dissolved oil vary widely among tests, especially considering differences among crude and refined oils in physical properties such as viscosity. The lack of tools to measure mixing energy is a major obstacle to standardizing mixing methods.

Experiments with dispersed oil require appropriate controls including dispersant-only to measure the toxicity of dispersant free in solution, and chemically dispersed nontoxic mineral oil to show whether dispersant contributes to observed oil toxicity. Adams et al. (2014a) found no effects on rainbow trout embryos of mineral oil CEWAF prepared with COREXIT 9500 (Nalco) and high energy mixing at a dispersant to oil ratio of 1:20. In contrast, mineral oil CEWAF prepared with low energy stirring at a ratio of 1:10 was toxic to fathead minnow embryos (Madison et al. 2015), suggesting a species difference in sensitivity to mineral oil or the incomplete mixing of dispersant with oil. Unmixed dispersant in CEWAF may reflect the wide array of reported mixing conditions (Supplemental Data, Table S1) and the high viscosity of some crude and heavy fuel oils. Research is needed to assess how methods for applying and mixing dispersant affect the efficiency of mixing and the concentrations of free dispersant in solution.

Research needs

1) Assess the effect of different mixing protocols reported in the literature on the composition and toxicity of WAF and CEWAF.
2) Develop methods or models to measure mixing energies for WAF and CEWAF solutions.
3) Establish approaches to optimize the addition of dispersants to oil and to reduce concentrations of free dispersant in CEWAF.

OIL STORAGE, HANDLING, AND WEATHERING

Experiments with weathered oil are difficult to interpret without detailed chemical characterization. The toxicity of naturally weathered oil is highly variable because weathering represents a continuum of change in chemical composition, not a binary either/or state. Artificial weathering to remove low molecular weight volatile compounds by aeration, heating (e.g., ASTM International 2018c; Method D2892-16, fractional distillation at <130 °C), or low temperature vacuum distillation (ASTM
Comparative with mixing vessels, the rapid loss of volatiles that cause acute lethality must be minimized by covering test containers to create a minimal head space and using only sufficient aeration to maintain 60% saturation (US Environmental Protection Agency 1996; Organisation for Economic Co-operation and Development 2018). However, data on biomass loading rates, water quality, photoperiod, and head space were not reported sufficiently often to collect statistics (Supplemental Data, Table S2), which introduces further uncertainties in test results. For example, CROSERF recommendations to conserve volatiles in acute lethality tests by sealing test vessels, restricting aeration, and not removing dead organisms may cause significant declines in dissolved oxygen concentrations. Most guidance documents recommend that oxygen be sustained at concentrations needed for growth and survival to avoid confounding test results (e.g., Environment Canada 2007).

Hydrocarbons are also removed from solution by other pathways including partitioning to the materials of test vessels, dead fish, uneaten food, fish waste, and biofilms that develop in response to accumulating waste. Hydrophobic compounds partition more readily to plastics and to silicone (polymethylsiloxane) sealant than to glass or stainless steel but many tests were conducted with plastics that may absorb hydrocarbons (Supplemental Data, Table S2). There is also a potential for microbial degradation of oil during prolonged mixing and settling of stock solutions and in static toxicity tests (CROSERF; Aurand and Coelho 2005). Although biodegradation may reduce hydrocarbon concentrations, some hydroxylated metabolites of polynuclear aromatic compounds are more toxic than the parent compound (Fallantaft et al. 2012). For instance, the embryotoxicity of an alkyl polynuclear aromatic compound (7-isopropyl-1-methylphenanthrene) increased when static test solutions were renewed daily without removing biofilms (P.V. Hodson, unpublished observation). Lighting that includes ultraviolet wavelengths can dramatically reduce toxicity because photolysis degrades PAC (Lee 2003). However, photo transformations of PAC that have accumulated in tissues of semi-transparent fish embryos may cause tissue necrosis and rapid mortality because reactive oxygen species are produced in vivo (Barron et al. 2005). Unless an oil toxicity test assesses the effects of photo transformations, lighting that emits ultraviolet wavelengths should be avoided.

Recommendation

Test conditions that affect steady-state concentrations of oil in test solutions should be optimized (e.g., Organisation for Economic Co-operation and Development 2018) to define standard conditions that produce accurate and repeatable test results.

CHARACTERIZING TEST SOLUTIONS

Without chemical analyses of test solutions, toxicity will be significantly underestimated when endpoints are calculated from nominal loadings or dilutions of WAF and CEWAF (Bejarano et al. 2014). Test solutions must be analyzed to

Recommendations

Standard methods should be followed for storing, mixing, and sampling oil under conditions that preserve its integrity (e.g., ASTM International 2018a, Method 4057-12; Aurand and Coelho 2005), as indicated by physical and chemical analysis when experiments begin, and after experiments that exceed 4 wk. Weathering during solution preparation should be assessed by chemical analysis or minimized by adjusting mixing conditions.

TEST CONDITIONS

For many oil toxicity tests, test conditions followed guidelines published by CROSERF or a national authority (e.g., Environment Canada 2007), modified to recognize regional characteristics (e.g., freshwater vs marine fish species). Despite these guidelines, critical test conditions that can bias toxicity tests appear to be ignored or not reported in many studies (Supplemental Data, Table S2) perhaps because national guidelines are generic and do not specifically address causes of rapidly changing oil concentrations in static and semi-static tests. To illustrate, guidelines may recommend that volumes of test solutions correspond to the size and biomass of test organisms. In the literature review, reported test volumes ranged from <1.0 mL (microplate wells for zebrafish embryos) to 100 L (juvenile fish; Supplemental Data, Table S2). Assuming that hydrophobic compounds will partition to the walls of test vessels, the proportional loss of oil from test solutions will increase with decreasing sizes of test vessels and increasing ratios of surface area to volume. Similarly, the loss of oil from test solutions will increase with biomass loading rates (total grams of fish per liter of test solution) because of bioaccumulation.

International 2018c; Method D1160-15, 200 °C) is not equivalent to natural weathering during an oil spill, which includes evaporation, water washing, biodegradation, and photodegradation. To understand the effects of weathering on toxicity, unweathered oil should be tested as a control.

Weathering also occurs during toxicity testing but few reports have detailed how test oils were stored, mixed, or repeatedly subsampled (Adams et al. 2017). Volatiles escape each time a container is opened, and losses may continue during storage if containers are not sealed and stored cold with a minimal head space (Aurand and Coelho 2005). Otherwise, the properties of oil may shift progressively from the start to the end of an experiment, with noticeable changes in the composition and toxicity of test solutions. For example, when exposed to air, diluted bitumen weatheres more rapidly than conventional crude oils as volatile diluents evaporate, causing rapid increases in viscosity and adhesiveness and reductions in buoyancy and dispersibility (Government of Canada 2013). During prolonged stirring of diluted bitumen and water to prepare WAF or CEWAF, weathering caused a noticeable slowing of stir bar speed, a reduced effectiveness of chemical dispersion (Madison et al. 2015), and possible changes in chronic toxicity; similar issues may occur with light shale oils.

Oil toxicity test methods improved—Environmental Toxicology and Chemistry, 2019,38:302–311 307

wileyonlinelibrary.com/ETC
describe gradients of oil concentration, changes in concentrations between solution renewals (static and semi-static regimes), and trends in concentration over the course of the experiment. Out of 144 studies surveyed, only 5% reported the full suite of hydrocarbon analyses recommended by CROSERF (total petroleum hydrocarbons [TPHs], volatile organic carbons [VOCs], and PAC; Aurand and Coelho 2005), and 16% did not include any of these recommendations (Adams et al. 2017). For acute toxicity tests, it is important to measure VOCs and low molecular weight (<C12) diaromatics (e.g., naphthalene, alkyl naphthalenes) and alkanes, compounds associated with acute lethality (Redman et al. 2017); chronic toxicity is associated with 3-5 ringed alkyl PAC (Hodson et al. 2007; Adams et al. 2014a; Bornstein et al. 2014).

Cost is a significant obstacle to analyses. A typical 96-h static daily renewal test could generate 30 to 40 samples to describe gradients of oil concentration and changes in concentration between solution renewals (0–24 h) and over the course of the experiment (96 h); replication would add more. Although TPH analyses are less expensive than PAC analyses, they are less informative about which analytes might be toxic.

Alternative, low-cost semi-quantitative techniques are available, specifically the measurement of TPH by fluorescence spectrometry (TPH-F; ~1% of the cost of polynuclear aromatic compound measurement), calibrated against known dilutions of oil (Adams et al.; Martin et al. 2014). As well, TPH-F analyses are rapid (minutes), allowing frequent sampling to describe the variance and temporal trends of oil concentrations and corrective feedback if needed. Although TPH-F analyses do not quantify individual hydrocarbons, fluorescence is strongly correlated to TPHs and the sum of all PAC concentrations (total PAC or TPAC) (Martin et al. 2014); thus toxicity data can be converted from TPH-F to TPH or TPAC by linear regressions when all 3 are measured in an affordable number of samples.

Most studies express chronic toxicity in terms of TPAC concentrations, which may include 50 or more analytes, including individual TPAC (e.g., unsubstituted chrysene) and families of alkyl PAC (e.g., C2-chrysenes; Forth et al. 2017b). However, the relationship between effects and measured TPAC concentrations will change if test solutions contain different arrays of PAC because the toxicity of petrogenic PAC varies widely with alkyl substitution and octanol–water partition coefficients (Kow; Hodson 2017). Comparisons of toxicity among oils or test conditions may also be confounded if concentrations of a large portion of the PAC analyzed are below the detection limits of the applied analytical method, which is often the case for chronic toxicity.

Chemical analyses are not without error. Only dissolved hydrocarbons appear to be bioavailable and toxic to some species (Carls et al. 2008; Redman et al. 2017), although oil droplets adhering to eggs of Atlantic haddock (Melanogrammus aeglefinus) coincided with a greater uptake of PAC, and a higher sensitivity to oil exposure than eggs of Atlantic cod (Gadus morhua; Sørhus et al. 2015; Sørensen et al. 2017). Oil droplets in test solutions will be combined with dissolved oil in total oil when water samples are extracted with solvents for TPH and PAC analyses, or when samples for TPH-F analyses are preserved with ethanol. Hence expressing toxicity in terms of total oil concentrations will underestimate or overestimate the toxicity of dissolved oil when droplets are present, depending on the species (Martin et al. 2014; Redman and Parkerton 2015). Few studies have discriminated particulate from dissolved oil, likely because of cost.

Chemical dispersion increases the concentrations of oil droplets in WAF and the partitioning of hydrocarbons from oil to water. The greater toxicity of CEWAF than WAF has contributed to conclusions that dispersants cause synergistic toxicity with oil (e.g., Rico-Martinez et al. 2013). Nevertheless, when toxicity was calculated from measured hydrocarbon concentrations, there was little difference in chronic toxicity between WAF and CEWAF (Aurand and Coelho 2005; Wu et al. 2012; Alsaadi et al. 2018) and no compelling evidence of synergism (Adams et al. 2014b). At the very low concentrations of oil causing chronic toxicity (0.2–10 μg/L of TPAC; Lee et al. 2015), most PAC are likely in the dissolved phase (Forth et al. 2017a).

To assess whether dispersants have contributed to CEWAF toxicity, concentrations of unique components such as dioctyl sodium sulfosuccinate (e.g., Place et al. 2016) should be measured, separating dissolved dispersant from dispersant sequestered into oil droplets. However, dispersants and their components are not routinely analyzed in test solutions and there appear to be no methods for distinguishing dissolved from sequestered dispersant.

Filtration of test solutions before solvent extraction can remove droplets before chemical analysis (Sandoval et al. 2017). Removal can be verified by measuring water insoluble components such as hopane, a key indicator of liquid-phase crude oil (Bennett et al. 1990), or changes in fluorescence spectra (Sandoval et al. 2017). However, oil retained on filters might still bias analyses if it is a source or sink for dissolved hydrocarbons during filtration (Organisation for Economic Co-operation and Development 2018); dissolved hydrocarbons may also partition to clean filters.

As an alternative, Redman and Parkerton (2015) recommend passive sampling devices to accumulate dissolved hydrocarbons from water, and oil partitioning models to estimate the bioavailable fraction from Kow of individual compounds. In combination with models of pharmacokinetics and toxic units, this biomimetic solid-phase microextraction of test solutions can reduce errors in estimated toxicity when oil concentrations change over time (Letinski et al. 2014; Bera et al. 2018). This approach has also shown that low molecular weight soluble hydrocarbons contribute more to acute lethality than larger compounds (Redman and Parkerton 2015). A similar approach has been applied to chronic toxicity using average acute to chronic ratios (Bera et al. 2018) but validation is needed with actual chronic toxicity tests.

Biomarkers of hydrocarbon accumulation can also be used to integrate temporal variations in oil concentrations. For example, the expression of cytochrome P4501A mRNA (cyp1a), and the subsequent synthesis and activity of CYP1A enzymes that oxygenate polynuclear aromatic compounds, increase with oil exposure by up to 100-fold compared with controls. In fish embryos exposed to diluted bitumen, the extent of cyp1a
expression was strongly correlated to the prevalence of deformities (Madison et al. 2017; Alsaadi et al. 2018), suggesting its use to indicate exposure to toxic concentrations of oil.

Data quality objectives and programs of quality assurance and quality control are essential components of analytical chemistry in support of toxicity testing. A quality assurance and quality control program should measure and report: 1) detection limits for each analyte or summed parameter (e.g., total polynuclear aromatic compounds), 2) percentage recoveries of spiked standard solutions, 3) performance on blind analyses of duplicate and reference samples, and 4) evidence that reported concentrations are within the linear range of the analytical method. Analytical laboratories should also participate in interlaboratory comparisons of performance. Unfortunately, few reports of oil toxicity describe a quality assurance and quality control program or provide details on performance, in part because there are no interlaboratory programs to disseminate reference oils and analytical standards.

**Recommendation**

Oil toxicity tests should be supported with sufficient analytical chemistry to describe gradients of test concentrations and daily and longer term changes in oil concentrations, including data quality programs to ensure the reliability of analytical data.

**Research needs**

1) Develop cost-effective methods for measuring particulate and dissolved oil and free and sequestered dispersant concentrations in toxicity test solutions.
2) Optimize alternative methods for characterizing oil exposure, including models and biomarkers to estimate tissue concentrations of hydrocarbons for different types of oil and test conditions.

**TOXICITY TEST REPORTS**

Supplemental Data, Tables S1 and S2 indicate that a significant weakness throughout the literature is a lack of detail on test characteristics that could affect measured oil toxicity (Adams et al. 2017). Studies that have claimed to follow specific guidelines (e.g., CROSERF) often deviate from recommended procedures without providing details or reasons for these changes and their implications on test outcomes. Some reports of toxicity measured under realistic site-specific oil spill conditions did not describe those conditions, how the test reproduced those conditions, what assumptions were inherent in the test design, and whether appropriate controls were included, particularly where oil was weathered or dispersants applied. Few reports described the physical and chemical characteristics of the oils as tested, and how the integrity of the oil sample was maintained during storage and handling. Similarly, many papers did not adequately document oil exposures including: 1) lists of analytes measured, 2) detection limits, 3) quality assurance and quality control data, 4) temporal trends in concentrations, 5) relationships between oil loadings and measured oil concentrations, and 6) how summed parameters such as TPAC were calculated. There was also little information on the origins, culture conditions, physiological status, size, feeding rates, mortality rates, or unusual behaviors of test organisms before and during testing; biomass loading rates per treatment; and materials and construction of test vessels. All these facts should be included in toxicity test reports, and the accessibility of journal supplemental data files removes any limits to providing details.

**Recommendation**

Checklists of data needed in oil toxicity test reports should be developed and used as a basis for deciding their acceptability for publication.

**DEVELOPING METHODS THROUGH COLLABORATION**

The suggestions in the present review for oil toxicity testing may not suit all practitioners. To translate these suggestions into widely accepted standard methods requires a collaborative process analogous to that used by CROSERF. If industry, government, and academia can achieve a consensus, it is more likely that the methods recommended will be applied broadly to promote comparability of data among experiments and a broader database for risk and impact assessments.

**Recommendation**

Industry, government, and university scientists should be engaged in an international, collaborative effort to improve and validate standard oil toxicity test methods.

**SUMMARY**

Too few studies of oil toxicity have applied standard or consistent protocols to facilitate comparisons of toxicity among oils and test conditions. Potential biases in test results are associated with variable exposure regimes that are not well characterized, inappropriate conditions for storing and sampling test oils, diverse methods for preparing solutions of oil in water, and test conditions that cause rapid changes in the composition and concentrations of test solutions. Some differences in toxicity among tests can be resolved by expressing toxicity as the measured concentrations of toxic components; however, chemical analyses are often too infrequent to characterize important changes in oil concentrations and composition throughout the test. The lack of practical methods to discriminate dissolved from particulate oil also creates biases in the estimated toxicity of test solutions. The overall effect of a diversity in test protocols and of the effects of methods on toxicity is a greater but unquantified uncertainty in toxicity benchmarks and ecological risk assessments. Research is needed to develop standard methods for unambiguous
measures of oil toxicity that will improve the understanding of why oil toxicity changes under site-specific conditions.

**Supplemental Data**—The Supplemental Data are available on the Wiley Online Library at DOI:10.1002/etc.4303.

**Acknowledgment**—The present report is based on a review of oil toxicity test methods sponsored by the National Contaminants Advisory Group of Fisheries and Oceans Canada (Adams et al. 2017). The literature review was funded by a contract to P.V. Hodson and R.S. Brown from the aforementioned agency (FP802-150296). The contract report was reviewed by 24 experts from government, universities, and the oil industry at Fisheries and Oceans Canada Science Advisory Workshop—Towards the Development of Toxicity Standard Methods to Evaluate Biological Effects of Heavy Oils on Aquatic Ecosystems—in Ottawa, Ontario, Canada, January 31 through February 2, 2017. We thank the above authorities for sharing their knowledge; and we are grateful to J. Leblanc and C. Lougheed of the Department of Fisheries and Oceans for their encouragement and support, and 2 anonymous reviewers who provided important comments that improved this report.

**Disclaimer**—All authors declare no conflicts of interest.

**Data Accessibility**—Data cited in this review are summarized from Adams et al. (2017), which is accessible at the URL provided in the references.

**REFERENCES**

Adams J, Bornstein JM, Munno K, Hollebone B, King RT, Brown RS, Hodson PV. 2014a. Identification of compounds in heavy fuel oil that are chronically toxic to rainbow trout embryos by effects-driven chemical fractionation. *Environ Toxicol Chem* 33:825–835.

Adams J, Sweeze M, Hodson PV. 2014b. Oil and oil dispersant do not cause synergistic toxicity to fish embryos. *Environ Toxicol Chem* 33:107–114.

Adams J, Charbonneau K, Tuor D, Brown RS, Hodson PV. 2017. Review of methods for measuring the toxicity to aquatic organisms of the water accommodated fraction (WAF) and chemically-enhanced water accommodated fraction (CEWAF) of petroleum. Research Document 2017/064, Department of Fisheries and Oceans, Ottawa, Ontario, Canada. [cited 2018 October 1]. Available from: http://www.dfo-mpo.gc.ca/csas-sccs/Publications/ResDocs-DoctRech/2017/2017_064-eng.html

Alsaadi FM, Madison BN, Brown RS, Hodson PV, Langlois VS. 2018. Morphological and molecular effects of two diluted bitumens on developing fathead minnow (Pimephales promelas). *Aquat Toxicol* 204:107–116.

Anderson JW, Neff JM, Cox BA, Tatem HE, Hightower GM. 1974. Characteristics of dispersions and water-soluble extracts of crude and refined oils and their toxicity to estuarine crustaceans and fish. *Mar Biol* 27:75–88.

ASTM International. 2018a. Standard practice for manual sampling of petroleum and petroleum products. Method D4057-12. Philadelphia, PA, USA. [cited 2018 October 1]. Available from: https://www.astm.org/search/fullsite-search.html?qury=D4057&toplevel=products-and-services&sublevel=standards-and-publications

ASTM International. 2018b. Standard test method for distillation of crude petroleum (15-theoretical plate column). Method D2892-18. Philadelphia, PA, USA. [cited 2018 October 1]. Available from: https://www.astm.org/search/fullsite-search.html?qury=D2892&toplevel=products-and-services&sublevel=standards-and-publications

ASTM International. 2018c. Standard test method for distillation of petroleum products at reduced pressure. Method D1160-18. Philadelphia, PA, USA. [cited 2018 October 1]. Available from: https://www.astm.org/search/fullsite-search.html?qury=D1160&toplevel=products-and-services&sublevel=standards-and-publications

Aurand D, Coelho G. 2005. Cooperative aquatic toxicity testing of dispersed oil and the Chemical Response to Oil Spills: Ecological Effects Research Forum (CROSERF). Technical Report. Ecosystem Management & Associates, Lusby, MD, USA.

Barron MG, Caruso MG, Short JW, Rice SW, Heintz RA, Rau M, Di Giulio R. 2005. Assessment of the phototoxicity of weathered Alaska North Slope crude oil to juvenile pink salmon. *Chemosphere* 60:105–110.

Barron MG, Ka’aihue L. 2003. Critical evaluation of CROSERF test methods for oil dispersant toxicity testing under subarctic conditions. *Mar Pollut Bull* 46:1191–1199.

Bejarano AC, Clark JR, Coelho GM. 2014. Issues and challenges with oil toxicity data and implications for their use in decision making: A quantitative review. *Environ Toxicol Chem* 33:732–742.

Benett D, Girling AE, Bounds A. 1990. Ecotoxicology of oil products: Preparation and characterisation of aqueous test media. *Chemosphere* 21:659–669.

Bera G, Parkerton T, Redman A, Turner NR, Renegar DA, Sericano JL, Knap AH. 2018. Passive dosing yields dissolved aqueous exposures of crude oil comparable to the CROSERF (Chemical Response to Oil Spills: Ecological Effects Research Forum) water accommodated fraction method. *Environ Toxicol Chem* 37:2810–2819.

Bornstein JM, Adams JE, Hollebone B, King T, Hodgson PV, Brown RS. 2014. Effects-driven chemical fractionation of heavy fuel oil to isolate compounds toxic to trout embryos. *Environ Toxicol Chem* 33:814–824.

Carls MG, Holland L, Larsen M, Collier TK, Scholz NL, Incardona JP. 2008. Fish embryos are damaged by dissolved PAHs, not oil particles. *Aquat Toxicol* 87:121–127.

Carls MG, Rice SD, Hose JE. 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 pallasi*). *Environ Toxicol Chem* 18:481–493.

Croce B, Stagg RM. 1997. Exposure of Atlantic salmon parr (*Salmo salar*) to a combination of resin acids and a water soluble fraction of diesel fuel oil: A model to investigate the chemical causes of pigmentated salmon syndrome. *Environ Toxicol Chem* 16:1921–1929.

Danion M, Le Floch S, Lamour F, Guyomarch J, Quentel C. 2011. Effects of oil toxicity on larval European sea bass (*Dicentrarchus labrax L.*). *Environ Toxicol Safe* 74:2167–2174.

Echos BS, Smith A, Gardinali P, Rand G. 2016. An evaluation of select test variables potentially affecting acute oil toxicity. *Arch Environ Contam Toxicol* 70:392–405.

Environment Canada. 2007. Biological test method: Acute lethality test using rainbow trout. Report EPS1/RM/09. Ottawa, Ontario, Canada.

Fallatafi S, Rantanen T, Brown RS, Sneickus V, Hodson PV. 2012. Toxicity of hydroxylated alkyl-phenanthenes to the early life stages of Japanese medaka (*Oryzias latipes*). *Aquat Toxicol* 106:107–124.

Forth HP, Mitchelmore CL, Morris JM, Lay CR, Lipton J. 2017a. Characterization of dissolved and particulate phases of water accommodated fractions used to conduct aquatic toxicity testing in support of the *Deepwater Horizon* natural resource damage assessment. *Environ Toxicol Chem* 36:1460–1472.

Forth HP, Mitchelmore CL, Morris JM, Lipton J. 2017b. Characterization of oil and water accommodated fractions used to conduct aquatic toxicity testing in support of the *Deepwater Horizon* oil spill natural resource damage assessment. *Environ Toxicol Chem* 36:1450–1459.

Government of Canada. 2013. Properties, composition and marine spill behaviour, fate and transport of two diluted bitumen products from the Canadian oil sands. Technical Report. Ottawa, Ontario, Canada.

Heintz RA, Short JW, Rice SD. 1999. Sensitivity of fish embryos to weathered crude oil: Part II. Incubating downstream from weathered Exxon Valdez crude oil caused increased mortality of pink salmon (*Oncorhynchus gorbuscha*) embryos. *Environ Toxicol Chem* 18:494–503.

Hodson PV. 2017. The toxicity to fish embryos of PAH in crude and refined oils. *Arch Environ Contam Toxicol* 73:12–18.

Hodson PV, Khan C, Saravanabavan G, Clark J, Brown RS, Hollebone B, Wang Z, Short J, Lee K, King T. 2007. Alky1 PAHs in crude oil cause chronic toxicity to early life stages of fish. *Proceedings, Thirtieth Arctic and Marine Oilspill Program (AMOP) Technical Seminar, June 5-7, 2007, Edmonton, Alberta*. Environment Canada, Ottawa, Ontario, pp 291–300.

© 2018 The Authors
Kiparissis Y, Akhtar P, Hodson PV, Brown RS. 2003. Partition-controlled delivery (PCD) of toxicants: A novel in vivo approach for embryotoxicity testing. Environ Sci Technol 37:2262–2266.

Lee RF. 2003. Photo-oxidation and photo-toxicity of crude and refined oils. Spill Sci Technol B 8:157–162.

Lee K, Boufadel M, Chen B, Foght J, Hodson PV, Swanson S, Venosa A. 2015. Expert Panel Report on the Behaviour and Environmental Impacts of Crude Oil Released into Aqueous Environments. Royal Society of Canada, Ottawa, ON. ISBN: 978-1-928140-02-3.

Letinski D, Parkerton T, Redman A, Manning R, Bragin G, Febbo E, Palandro D, Nedwed T. 2014. Use of passive samplers for improving oil toxicity and spill effects assessment. Mar Pollut Bull 86:274–282.

Lin H, Morandi GD, Brown RS, Snieckus V, Rantanen T, Jørgensen KB, Hodson PV. 2015. Quantitative structure-activity relationships for chronic toxicity of alkyl-chrysenes and alkyl-benz[a]anthracenes to Japanese medaka embryos (Oryzias latipes). Aquat Toxicol 159:109–118.

Madison BN, Hodson PV, Langlois VS. 2015. Diluted bitumen causes deformities and molecular responses indicative of oxidative stress in Japanese medaka embryos. Aquat Toxicol 165:222–230.

Madison BN, Hodson PV, Langlois VS. 2017. Cold Lake Blend diluted bitumen toxicity to the early development of Japanese medaka. Environ Pollut 225:579–586.

Martin JD, Adams J, Hollebone BP, King T, Brown RS, Hodson PV. 2014. Chronic toxicity of heavy fuel oils to fish embryos using multiple exposure scenarios. Environ Toxicol Chem 33:677–687.

Nordtug T, Olsen AJ, Altin D, Meier S, Overrein I, Hansen BH, Johansen O. 2011. Method for generating parameterized ecotoxicity data of dispersed oil for use in environmental modelling. Mar Pollut Bull 62:2106–2113. Organisation for Economic Co-operation and Development. 2018. Guidance document on aqueous-phase aquatic toxicity testing of difficult test chemicals. Series on Testing and Assessment, No. 23, 2nd ed. ENV/JM/MONO(2000)6. Paris, France.

Pan Z, Zhao L, Boufadel MC, King T, Robinson B, Conney R, Lee K. 2017. Impact of mixing time and the dispersion effectiveness and droplets size of oil. Chemosphere 166:246–254.

Place BJ, Perkins MJ, Sinclair E, Barsamian AL, Blakemore PR, Field JA. 2016. Trace analysis of surfactants in Corexit oil dispersant formulations and seawater. Deep Sea Res Part 2 Top Stud Oceanogr 129:273–281.

Redman AD, Butler JD, Letinski DJ, Parkerton TF. 2017. Investigating the role of dissolved and droplet oil on aquatic toxicity using dispersed and passive dosing systems. Environ Toxicol Chem 36:1020–1028.

Redman AD, Parkerton TF. 2015. Guidance for improving comparability and relevance of oil toxicity tests. Mar Pollut Bull 98:156–170.

Rico-Martinez R, Snell TW, Shearer TL. 2013. Synergistic toxicity of Macondo crude oil and dispersant Corexit 9500A to the Brachionus plicatilis species complex (Rotifera). Environ Pollut 173:5–10.

Sandoval K, Ding Y, Gardinali P. 2017. Characterization and environmental relevance of oil water preparations of fresh and weathered MC-252 Macondo oils used in toxicology testing. Sci Total Environ 576:118–128.

Sørensen L, Sørhus E, Nordtug T, Incardona JP, Linbo TL, Giovanetti L, Karlsen O., Meier S. 2017. Oil droplet fouling and differential tocostokinetics of polycyclic aromatic hydrocarbons in embryos of Atlantic haddock and cod. PLoS One 12:e0180048.

Sprague JB. 1969. Measurement of pollutant toxicity to fish I. Bioassay methods for acute toxicity. Water Res 3:793–821.

US Environmental Protection Agency. 1996. Method 850.1075. Ecological effects test guidelines. Fish acute toxicity test, freshwater and marine. Report EPA 712–C–96–118. Washington, DC, United States of America.

Wu D, Wang Z, Hollebone B, McIntosh S, King T, Hodson PV. 2012. Comparative toxicity of four chemically dispersed and undispersed crude oils to rainbow trout embryos. Environ Toxicol Chem 31:754–765.

Zhao L, Bing W, Armenante PM, Robyn C, Boufadel MC. 2016. Characterization of turbulent properties in the EPA baffled flask for dispersion effectiveness testing. J Environ Eng 142:4015044.