8-1-2023

Improving the design and conduct of aquatic toxicity studies with oils based on 20 years of CROSERF experience

WA Stubblefield
Oregon State University

M Barron
United States Environmental Protection Agency

G Bragin
ExxonMobil Biomedical Sciences

ME DeLorenzo
National Oceanic and Atmospheric Administration (NOAA)

B de Jourdan
Huntsman Marine Science Centre

See next page for additional authors

Authors
B Echols
Environmental Toxicology Associates

DP French-McCay
RPS Group

P Jackman
Environment and Climate Change Canada

JR Loughery
Huntsman Marine Science Centre

TF Parkerton
EnviSci Consulting

D. Abigail Renegar
Nova Southeastern University, drenegar@nova.edu

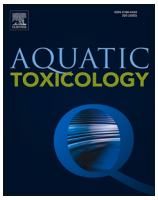
JL Rodriguez-Gil
International Institute for Sustainable Development Experimental Lakes Area (IISD-ELA)

Aquatic Toxicology 261 (2023) 106579

Contents lists available at ScienceDirect

Aquatic Toxicology

journal homepage: www.elsevier.com/locate/aqtox



Improving the design and conduct of aquatic toxicity studies with oils based on 20 years of CROSERF experience

WA Stubblefield [a,*], M Barron [b], G Bragin [c], ME DeLorenzo [d], B de Jourdan [e], B Echols [f], DP French-McCay [g], P Jackman [h], JR Loughery [e], TF Parkerton [i], DA Renegar [j], JL Rodriguez-Gil [k]

[a] Oregon State University, USA
[b] United States Environmental Protection Agency (retired), USA
[c] ExxonMobil Biomedical Sciences, Inc., USA
[d] National Oceanic and Atmospheric Administration (NOAA), USA
[e] Huntsman Marine Science Centre, St. Andrews, New Brunswick, Canada
[f] Environmental Toxicology Associates LLC, USA
[g] RPS Group, USA
[h] Environment and Climate Change Canada (retired), Canada
[i] EnviSci Consulting, LLC, USA
[j] Nova Southeastern University, USA
[k] International Institute for Sustainable Development Experimental Lakes Area (IISD-ELA), Canada

ARTICLE INFO

Keywords:
CROSERF
Aquatic toxicity test methods
Aquatic toxicology

ABSTRACT

Laboratory toxicity testing is a key tool used in oil spill science, spill effects assessment, and mitigation strategy decisions to minimize environmental impacts. A major consideration in oil toxicity testing is how to replicate real-world spill conditions, oil types, weathering states, receptor organisms, and modifying environmental factors under laboratory conditions. Oils and petroleum-derived products are comprised of thousands of compounds with different physicochemical and toxicological properties, and this leads to challenges in conducting and interpreting oil toxicity studies. Experimental methods used to mix oils with aqueous test media have been shown to influence the aqueous-phase hydrocarbon composition and concentrations, hydrocarbon phase distribution (i.e., dissolved phase versus in oil droplets), and the stability of oil:water solutions which, in turn, influence the bioavailability and toxicity of the oil containing media. Studies have shown that differences in experimental methods can lead to divergent test results. Therefore, it is imperative to standardize the methods used to prepare oil:water solutions in order to improve the realism and comparability of laboratory tests. The CROSERF methodology, originally published in 2005, was developed as a standardized method to prepare oil:water solutions for testing and evaluating dispersants and dispersed oil. However, it was found equally applicable for use in testing oil-derived petroleum substances. The goals of the current effort were to: (1) build upon two decades of experience to update existing CROSERF guidance for conducting aquatic toxicity tests and (2) to improve the design of laboratory toxicity studies for use in hazard evaluation and development of quantitative effects models that can then be applied in spill assessment. Key experimental design considerations discussed include species selection (standard vs field collected), test substance (single compound vs whole oil), exposure regime (static vs flow-through) and duration, exposure metrics, toxicity endpoints, and quality assurance and control.

1. Introduction

Toxicity testing is a key aspect of oil spill science and decision making and in the development of strategies used to assess and mitigate environmental impacts. Toxicity tests are performed to address questions pertaining to the hazards, risks, and impacts of oil spilled in the aquatic environment and how temporal and spatial changes, due to environmental fate processes, can result in differing environmental effects. Decisions regarding spill-response mitigation procedures (e.g., dispersant use, beach cleaning) are in part based on our understanding

* Corresponding author.
E-mail address: bill.stubblefield@oregonstate.edu (W. Stubblefield).

https://doi.org/10.1016/j.aquatox.2023.106579
Received 28 December 2022; Received in revised form 31 March 2023; Accepted 3 April 2023
Available online 19 May 2023

of the environmental fate and toxicological effects attributable to spilled oil, and these are most often derived from laboratory-based testing. However, there are several other questions commonly addressed using empirical data developed from laboratory investigations, and some of these are described in Table 1.

A major challenge with respect to oil toxicity testing is that it is impossible to replicate, under laboratory conditions, the range of spill conditions, oil types, weathering states, receptor organisms, and modifying environmental factors that exist in actual oil spills. Further, oil concentrations in many spill situations can be substantially lower than those commonly used in toxicity tests (Neff and Stubblefield 1995; Echols et al., 2016). Petroleum and petroleum-derived products are complex mixtures composed of hydrocarbons (and related compounds) dictated chiefly by the geological source of crude oils and the processes used in the production of the refined petroleum products. Most components are unknown or uncharacterized and the current state-of-the-science is insufficient to permit a full characterization of all mixture components (Dettman et al., 2023). Once an oil enters the environment, its composition, distribution, and fate are dictated by properties such as volatility, solubility, and changes due to degradative processes such as photolysis, hydrolysis, and biodegradation, resulting in a hydrocarbon mixture that no longer resembles its pre-spill composition. Toxicity tests can be conducted with parent oils, field-collected samples, or natural or artificially weathered oil samples; however, all of these are typically representative of the composition of the hydrocarbon mixture and its toxicity only at a single location and point in time. Prediction and quantification of environmental impacts requires a quantitative characterization of a hydrocarbon mixture in an environmental sample, an understanding of the toxicity of the components, and the ability to relate composition to toxicity for a given duration of exposure. This understanding is critical to inform decision-making for oil spill response activities and to assess oil-associated environmental risks and impacts.

The goals of this article are to: (1) build upon two decades of experience to update existing Chemical Response to Oil Spills: Ecological Effects Research Forum (CROSERF) guidance (Aurand and Coelho 2005) for conducting aquatic toxicity tests and (2) to improve the design of laboratory toxicity studies for use in hazard evaluation and development of quantitative effects models that can then be applied in spill assessment. Key experimental design considerations discussed include species selection (standard vs field collected), test substance (single compound vs whole oil), exposure regime (static vs flow-through) and duration, exposure metrics, toxicity endpoints, and quality assurance and control. Our focus is on whole organism laboratory exposures to crude oil and related petroleum substance alone or in combination with spill response agents (SRAs) and excludes field exposures and sub-organismal or population level responses. These objectives are aligned with recent recommendations for improving oil toxicity tests (National Academies of Sciences and Medicine 2020). Greater standardization of laboratory studies on oil product and response agent toxicity will further advance the utility of study results in oil spill response, assessment, and predictive model development and application.

Table 1
Common objectives for conducting aquatic toxicity tests with oil and/or spill response agents.

Rationales for Oil Aquatic Toxicity Testing
Investigate the relative sensitivity of different organisms, life-stages or effect endpoints following exposure to a given substance
Compare the hazard of different oils or oil-derived materials using the same organism, experimental conditions, and effect endpoint(s)
Compare the hazard of oils at different weathering states using the same organism, experimental conditions, and effect endpoint(s)
Provide information to inform dispersant-use decisions in spill response planning
Determine acute to chronic ratios for a given organism and substance
Quantify the time-dependence of observed toxicity
Elucidate underlying mechanisms of observed toxicity
Quantify the role of modulating environmental factors (e.g., temperature, salinity, UV light, pressure, dissolved or particulate organic carbon) on observed toxicity
Evaluate the potential toxicity of oil degradation by-products
Evaluate the potential for latent effects after exposure is terminated
Provide input to toxicity models using single compound tests
Assess if toxicity models are predictive of observed toxicity

1.1. CROSERF method

Oils and related petroleum substances are complex mixtures comprised of thousands of compounds with widely different physical-chemical and toxicological properties; this leads to challenges conducting and interpreting oil toxicity studies. The experimental methods used to mix oils with water to produce aqueous test media can influence the component composition and concentrations, hydrocarbon phase distribution (i.e., dissolved phase versus in oil droplets), and the stability of oil:water solutions which, in turn, influence the bioavailability and subsequent toxicity of the oil containing media (Parkerton et al., 2023a, Redman et al., 2012). Prior to the development of the CROSERF method, tests were typically conducted following the method first published by Anderson et al. (1974). This involved the mixing of oil and water at a ratio of 1:9 in a 20 L glass vessel; the solution was slowly stirred for 20 h at room temperature, following which the water and oil phases were allowed to separate over a 1–6 h period and the water phase was then decanted and used for subsequent testing. The resulting oil:water mixture was termed a water accommodated fraction (WAF) or water soluble fraction (WSF). The WAF term has been more universally adopted because it encompassed the potential for droplet oil in the water-oil mixture.

The CROSERF methodology (Aurand and Coelho 2005) was developed specifically to address concerns about the adequacy of laboratory toxicity data used to characterize the effects of dispersants employed in marine oil spill response. The goal of the CROSERF project was two-fold: (1) provide a mechanism to screen potential chemical response agents, i.e., dispersants, and (2) help estimate the potential ecological effects of the use of dispersants. Early in the project, it was noted that there were difficulties working with oils and dispersed oils in aqueous solutions and that small differences in experimental protocols (e.g., mixing times, mixing energy, solution preparation method) could lead to large differences in test results, and differences in exposure metrics (e.g., loading,%WAF, TPH, TPAH, see Section 3.6), could lead to large differences in interpretation (see also Parkerton et al., 2023b). It therefore became an imperative to develop standardized methods to prepare oil: water solutions to improve the realism and comparability of laboratory toxicity tests. Although the CROSERF method was developed with a focus on dispersants and dispersed oil, the method was found applicable for use in testing oil-derived petroleum substances as well.

Over the decades since the publication of the original CROSERF methodology, there has been a proliferation of experimental methods and exposure metrics used in oil toxicity studies (Adams et al., 2017). The adoption of multiple, varied testing approaches has often led investigators to divergent conclusions (i.e., differing interpretations of study data and resulting implications) that can confuse stakeholders and potentially impede sound decision-making. Examples of investigations where the experimental approach or exposure metric used altered conclusions on observed oil toxicity include the role of mixing energy (Parkerton et al., 2023a), oil weathering state (Bobra et al., 1983; Di Toro et al. 2007), dispersant addition (National Academies of Sciences, Engineering 2020) or modifying factors such as the influence of ultraviolet (UV) light (Alloy et al., 2023). As a result, the adoption of a unifying conceptual model framework to guide the conduct and interpretation of laboratory oil toxicity tests is critically needed to advance the current state-of-the-science for oil effects assessment. That said, it is recognized that the experimental methods being employed must provide data useful for addressing specific questions. For that reason, it is not possible for the authors to provide specific recommendations for experimental methods to be used in conducting future oil testing. It is

recommend that investigators consider the approaches and concerns described herein and adopt those methods best suited to address the questions being posed.

1.2. Conceptual model framework for addressing mixtures

Foster et al. (2005) proposed a five-stage strategy for conducting environmental exposure assessments of mixtures, including (1) determination of mixture composition, (2) selection of component groups within the mixture, (3) compilation of relevant fate property data for each group, (4) assessment of environmental fate of each group, and (5) assessment of environmental exposure to each group and to the mixture as a whole. This strategy has been implemented in oil spill fate and exposure modeling, such that (1) available composition data for the oil are evaluated; (2) the oil composition is simulated using a number of pseudo-components (groups of compounds or "hydrocarbon blocks" with similar physical-chemical properties); (3) the physical-chemical properties and degradation rates of the pseudo-components are defined; (4) the fate of each pseudo-component is modelled in time and space for a given spill release scenario; and (5) the resulting environmental exposure for each of the pseudo-components and overall oil is calculated over space and time, as recently illustrated by French-McCay et al. (2018, 2021). These concepts can be extended to inform the conduct (this paper) and interpretation (Parkerton et al., 2023b) of oil toxicity tests. In addition to understanding oil composition, partitioning of components (i.e., in dissolved vs droplet form) following dosing and test organism exposure, key aspects of such a framework include accounting for the phase distribution and concentration of all hazardous components that may be present in laboratory test exposure media and how these components interact or combine to determine observed oil toxicity.

2. The aquatic hazard assessment framework for testing oil and oil-derived solutions

The evaluation of oils and their potential effects on environmental receptors is challenging for a variety of reasons, including the undefined composition of the hydrocarbon mixture, temporal and spatial changes in the composition of the oil/water mixture due to physicochemical properties and environmental fate processes, and the presence of dissolved components and undissolved oil droplets in the oil:water mixture. Four key issues have consistently occurred in oil toxicity testing, leading to difficulty in comparing the data or interpreting the results:

1) The first issue is how to mix oil and water; this aspect is addressed in the "media preparation" paper by Parkerton et al. (2023a). However, in this paper, emphasis is placed on how the objective of the study must drive the manner in which exposure solutions are prepared.
2) The second issue is how to standardize oil testing for hazard and impact assessment to allow for reproducibility of test results and comparability across oil products.
3) The third issue is how to take data from laboratory-based toxicity experiments and relate them to actual oil spill incidents. To address this issue, the use of biological effects models is recommended, as described in French-McCay et al. (2023), and guidance is provided on how to generate data that will be of use in developing models to predict biological effects based on oil type, weathering state, location, and exposure duration.
4) The fourth issue is that oils are mixtures and therefore it is difficult, if not impossible, to ascribe toxicity to the individual components when tested as a mixture. To address this issue, it is recommended that a reductionist approach is used to tease apart the contribution of individual components to toxicity via single compound experiments, followed by an assessment of component interactive effects.

When considered together, these four issues form the basis for the proposed framework for advancing the use of laboratory-based toxicity tests in combination with quantitative effects modeling of physically and chemically dispersed oils (Fig. 1).

This framework represents an experimental progression of approaches starting at the least complex, i.e., assessment of individual chemicals, and progressing toward the most complex, i.e., evaluating complex weathered and chemically dispersed oils. No "one size fits all" approach is sufficiently inclusive to address all applications; therefore, a variety of tools and approaches are required. The application of this framework for design of toxicity tests is explained in the sections that follow. The proposed framework is consistent with recently suggested approaches for the risk assessment of substances of unknown or variable composition, complex reaction products, or biological materials (UVCBs) (Salvito et al., 2020)

A thorough understanding of the intended use of the data is critical to the selection of the empirical test procedures to be employed (Table 1). Test species selection, test duration (i.e., acute or chronic), monitored effect endpoints, statistical design considerations, analytical chemistry sampling needs, and other aspects must be considered in deciding on an experimental design and the procedures to be used in test media preparation. The sections that follow address considerations for three testing approaches of increasing complexity.

2.1. Individual hydrocarbons

The foundation for hazard assessment of petroleum mixtures relies on understanding the hazards of the mixture components. As depicted in Fig. 1, the complexity of the hydrocarbon mixture will dictate the methods needed for conducting toxicity tests. For tests with single compounds, i.e., the more water-soluble hydrocarbon oil components (e.g., mono-aromatics, naphthalenes) or neat spill response agents, direct dosing of test substances can be used. For less-soluble hydrocarbons or complex mixtures, WAF preparation or passive dosing methods are preferred to allow better controlled exposures and more accurate toxicity test endpoints (Parkerton et al., 2023a).

The fundamental challenge in interpreting the toxicity of hydrocarbon mixtures is that the toxic properties of hydrocarbons vary widely. This means that simple methods that have been used historically, e.g., summing the concentrations of the individual PAH components, will not accurately reflect the toxicity of the complete mixture under all conditions because it assumes all the compounds are present in the same proportions as was tested when deriving the compared effects concentrations. Alternatively, the use of toxic units (TUs) (Hermans and Leeuwangh, 1982; Peterson, 1994) as the metric for expressing the toxicity of mixtures of oil-related components or mixtures of hydrocarbons has been demonstrated as an effective method of normalizing toxicity data across different sources and different species. The toxicity of the mixture depends on hydrocarbon composition, dissolved concentrations of the individual components, and the relative toxicity of the individual components. This concept forms the basis for modeling approaches such as the Target Lipid Model (TLM; Di Toro and Mcgrath 2000), OILTOXEX (French-McCay 2002a) and PETROTOX (Redman et al., 2012b).

Since it is impossible to conduct toxicity tests on thousands of individual compounds that comprise crude oil and petroleum-derived products, toxicity data are needed and should be available for selected "surrogate" hydrocarbons that can be used to develop quantitative structure activity relationships (QSARs) that will allow prediction of toxicity for untested compounds. Since laboratory assays conducted with whole oil are not representative of actual oil spill environmental conditions, a reductionist approach is necessary to use the test results from single hydrocarbon tests to develop and employ QSARs in additive models (i.e., summed TUs) to predict mixture effects. Oil toxicity tests support calibration and validation of toxicity models which predict environmentally realistic exposures (Hodson et al., 2019; National Academies of Sciences and Medicine 2020).

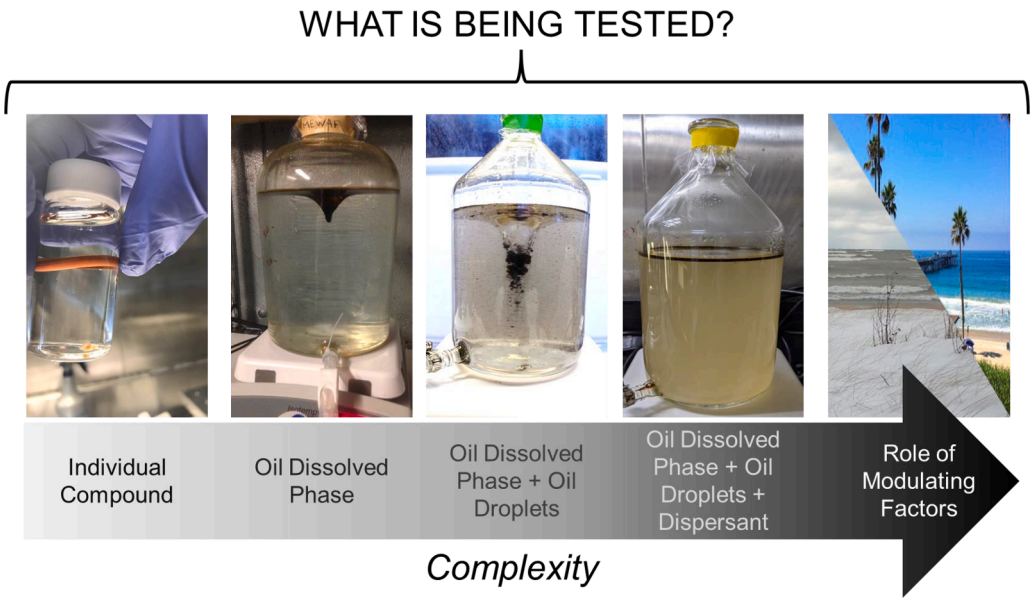


Fig. 1. Examples of increasingly complex approaches for testing crude oils, petroleum-derived products, and spill response agents. Depicted approaches become increasingly complex starting with testing of single compounds moving toward the most complex, i.e., whole oils with droplets and the presence of dispersant chemicals.

Equally important is an understanding of the range of sensitivity among species to individual compounds. Current methods used to protect aquatic communities involve the development of species sensitivity distributions (SSDs). An SSD is a probability distribution of acute or chronic hazard data collected for a given substance with different test species. These SSDs are widely used in decision-making to quantify interspecies sensitivity differences and this information is used in the derivation of protective hazard concentrations (i.e., the hazardous concentration for 5% of species, HC_5) intended to protect 95% of the exposed species (Barron et al., 2013; Bejarano 2018; Langdon and Rand 2018). While many standard test species have been included when deriving existing QSARs, potentially sensitive species and life stages that are difficult to study may or may not be characterized by available QSARs and included in SSDs.

2.2. Whole oil testing

2.2.1. Dissolved oil phase testing

Dosing of whole oil traditionally relies on preparation of a WAF, and the various methods for preparing WAFs are discussed in Parkerton et al. (2023a). One objective of whole-oil tests is to quantify the inherent toxicity of the oil product. Critical aspects of these tests are to provide data needed to make this characterization, i.e., the amount or loading of oil causing the adverse effect investigated (e.g., loading rate of test material that is expected to be lethal to 50% of a representative population, LL50), the time course of adverse population-relevant effects (e.g., survival, growth/development, reproduction), the shape or steepness of the concentration-response relationship and the asymptotic or incipient effect endpoint (Buikema et al., 1982; Sprague 1969) for the conditions tested.

In addition, whole oil bioassays using WAFs or other preparations exposing organisms to dissolved components are useful and important for validating the predictions of toxicity models based on additive effects of single compounds. Effect endpoints from WAF and passive dosing studies, such as the LC50 or EC50 (median lethal concentration or median effect concentration), are affected by the proportions of the subset of quantified hydrocarbons in the mixture and the duration of exposure to these hydrocarbons in the test. Toxicity tests employing varying exposure times and modifying factors, which are performed according to the criteria laid out herein, provide needed validation data for aquatic toxicity models. In such studies, it is important to account for the phase distribution and concentration of all hazardous components that may be present in laboratory test exposure media and how these components combine to determine observed oil toxicity.

2.2.2. Droplets and spill response agent (SRA) exposures

When preparing exposure media with significant droplet contribution, either from high-energy mixing (HEWAF) or the addition of a chemical dispersant (CEWAF), the choice of treatment preparation method greatly impacts the experimental objectives that can be addressed. Two approaches are commonly used for preparing treatments for evaluating concentration-response relationships. The first approach, referred to as variable loading, is based on the original CROSERF protocol and involves preparing individual WAFs at different oil loadings for each treatment (Singer et al., 2000). For tests performed using the variable loading approach, dissolved phase composition varies with oil loading thereby providing a range of oil profiles that is intended to represent various spill exposures in the field rather than evaluating a single, constant composition. For example, as the oil loading used to prepare the WAF stock increases, the dissolved phase component composition becomes increasingly enriched in the more water-soluble oil compounds, e.g., monoaromatiacs, naphthalenes). As a result, it is necessary to chemically characterize each exposure solution. In the case of WAFs prepared with addition of a chemical dispersant (CEWAF) or high energy mixing (HEWAF), test media will include both droplet and dissolved phases. Upon transfer of oil dosed media to exposure chambers with test organisms, dissolved phase exposures are expected to be maintained for CEWAF and HEWAF tests since the droplets that are present can serve as a reservoir to buffer potential losses due to volatilization, sorption, degradation or test organism uptake and metabolism. Thus, somewhat higher toxicity may be observed in variable loading tests where oil droplets are present versus test media preparations where droplet exposures are minimized or excluded (e.g., LEWAF (Low Energy Water Accommodated Fraction), MEWAF (Medium Energy Water Accommodated Fraction), passive dosing). However, droplet exposures are expected to be unstable, lead to direct exposure of elevated concentrations of low solubility oil constituents that may be unrepresentative of the field and cause physical effects, e.g., entrapment of test organisms in the rising oil at the surface of the test chambers. Another challenge is use of routine analytical methods for analyzing test media

with droplets that do not distinguish dissolved from particulate phase components.

The second approach, referred to as variable dilution, is based on a modification of the CROSERF protocol and involves preparing a WAF stock at a single oil loading and then making individual dilutions[1] of this stock as test treatments. Providing oil droplet concentrations and losses are negligible, the dissolved phase composition is constant across dilutions and depends on the oil loading selected to prepare the WAF stock. As a result, it may only be necessary to conduct a "full" chemical characterization of the stock solution, and then apply the appropriate dilution factor to calculate the concentration in the remaining exposure solutions. However, analysis of all exposure solutions using less costly techniques, such as fluorescence, is recommended to confirm proper dilution of prepared exposure solutions (see Dettman et al., 2023).

The nuances, advantages, and limitations of variable loading and variable dilution tests are described in Parkerton et al. (2023a) and are briefly summarized in Table 2. In conducting any test, one must consider the experimental objectives and these should dictate the best suited experimental approach.

The role of droplets in modulating dissolved phase exposures is evident when comparing CEWAF toxicity results for the same oil with different dispersants of comparable toxicity under a variable dilution testing scheme. At equal treatment dilutions, the toxic units will be greater for dispersants that are more effective at dispersing the oil. In other words, the more effective the dispersant the higher the dissolved oil exposures and resulting toxicity. However, if dispersant selection is decided on the basis of toxicity of oil+disperant, ineffective dispersants will be preferred thereby undermining efforts to mitigate impacts during spill response. Thus, common criteria used for dispersant selection based on effectiveness and toxicity are in obvious conflict. This recognition has led to revised selection criteria for dispersants to be based on the intrinsic toxicity of dispersant alone rather than the combined toxicity of oil+dispersant (Fieldhouse et al., 2019; Walton et al., 2021).

Testing of dispersant and other spill response agents (SRA; e.g., herders, demulsifiers, surface cleaners, sorbents, degreasers etc.) requires special consideration of the unique physical and chemical properties of the products as well as their intended usage rates and locations. Current unknowns and challenges with performing SRA-only exposures include the fact that they are mixtures which often contain poorly soluble/non-miscible components (e.g., herders stay on the surface), and that analytical techniques for exposure confirmation are limited (Dettman et al., 2023). Standardization of SRA testing will result in more comparable data and aid in the selection of the least toxic products with the greatest efficiency for spill response.

A final caution in conducting toxicity studies that investigate the effects of oil droplets or SRA is that test media that include such exposures are more prone to water quality changes that can confound test interpretation. Elevated concentrations of droplets or SRA can increase oxygen demand and cause unacceptable changes in dissolved oxygen levels needed to support healthy test organisms. Such water quality impacts may not be evident in controls or even treatments that include only dissolved phase exposures. Thus, it is critical to document acceptable water quality across treatments to ensure that study conclusions are reliable (Bejarano et al., 2023). Further, both oil droplets and SRA could lead to changes in the properties of the air-water interface, which can lead to physical effects (e.g., entrapment, Black et al., 2021) on the test organisms. If observed, these physical effects must be recorded and clearly distinguished from the effects resulting from the in-water exposure.

[1] It is recognized that a dilution series can be prepared by diluting an aliquot of a stock WAF for each exposure or by serially diluting a series of prepared solutions. Due to the importance of the assumption of homogeneity in solutions, it is not recommended to prepare exposure solutions using a serial dilution approach.

2.3. Incorporating modulating factors

2.3.1. Biotic and abiotic factors

The last aspect of Fig. 1 addresses factors that can influence either exposure or hazard concentrations. Biotic factors such as organism weight and age are key determinants of many allometric physiological functions, such as ingestion and respiration rates, and these may affect test substance uptake and influence exposure and effect concentrations. Abiotic conditions (e.g., pH, light, salinity, temperature, dissolved oxygen, hydrostatic pressure, particulate and dissolved organic carbon concentrations) can alter the physicochemical properties of the oil, leading to changes in bioavailability, as well as influence the physiological condition of aquatic organisms. Studies that incorporate abiotic variables as co-stressors can inform toxicity assessments across a wider range of environmental conditions.

A key recommendation when investigating the effects of modulating factors is to ensure that appropriate controls are included to ensure test organisms are healthy across the range of modulating factors being investigated and that acceptable water quality is documented in all treatments included in the study design (Bejarano et al., 2023). In designing an oil experiment with modifying factors, the abiotic conditions tested should be within the range of tolerance for the test organisms and within the normal range of climatic conditions for their habitat. Control survival should not be significantly affected by abiotic conditions. Acclimation to abiotic test conditions is an important experimental design consideration. The acclimation plan should be adjusted to meet the needs of the specific test organism. When modifying pH, the type of adjustment should be considered. In estuarine waters, hypercapnia (low pH as a result of elevated environmental carbon dioxide) can be defined as 1–2% CO_2 or a pH of 6.9–7.0, and the pH can be adjusted by mixing in gasses (oxygen, nitrogen, carbon dioxide) at different proportions to achieve different pH treatments. Oceanic acidification could be represented by adjustments to parameters in the carbonate system, including carbonate concentration, aragonite saturation, and dissolved inorganic carbon. For all exposures using environmental conditions as co-stressors, frequent measurement of all abiotic conditions is recommended throughout the exposure.

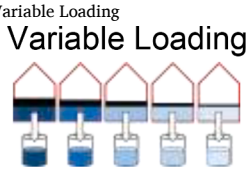
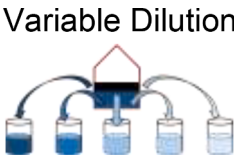
Ultraviolet (UV) light is a particularly important modifying factor of oil toxicity, with significant photoinduced toxicity documented for many marine and estuarine species (Alloy et al., 2023). To test without the influence of UV light, it is imperative that the oil preparations be made in the dark and that the oil exposures are conducted under standard laboratory fluorescent lighting. For detailed recommendations on incorporating UV light into oil exposures, readers are referred to Alloy et al. (2023).

The natural environment is highly variable, and every oil spill will coincide with a unique set of biotic and abiotic conditions. Characterizing the toxicity of oil under different exposure conditions will improve our understanding of the environmental impacts under various spill scenarios. An ideal approach would include a clearly defined spill scenario, development of a site-specific conceptual model, and adopting test substances, biotic aspects and abiotic parameters identified in the conceptual model to address the most relevant testing conditions and modifiers. Site-specific abiotic conditions may dictate implementation of unique site-specific experimental methods that may vary from "standard" methods but are necessary to produce meaningful and interpretable data.

2.3.2. Exposure dynamics

Exposure duration is a significant factor in observed toxicity, as shorter exposures (hours as opposed to days or longer) result in higher effect levels (less toxic). Similarly, constant exposures produce lower effect levels than spiked tests for the same average concentration and exposure duration (Aurand and Coelho 2005; Bejarano et al., 2014). Typical acute exposure times range from hours to days, with chronic exposures on the order of weeks or longer; this will vary based on the life

Table 2
Advantages and limitations of variable loading compared to variable dilution.

Solution Preparation Method	Example Objectives	Advantages	Limitations
Variable Loading 	• Hazard assessment of an oil • Compare physically and chemically dispersed oil • Compare species or life stages sensitivity	Each treatment is created individually and therefore better represent the range of potential compositions occurring in the field during a spill	Increased analytical costs associated with measuring each treatment created. Hydrocarbon compositional differences across exposure concentrations violate the assumption of consistency among test solutions, i.e., exposure solutions will have different composition not just concentrations.
Variable Dilution 	• Compare sensitivity of different species or life stages • Validate biological effects models	Reduced analytical costs as the treatment compositions can be calculated from the stock WAF composition (; Forth et al., 2017)	When used with higher loadings of fresh oil, mono aromatic and naphthalene components become more dominant in the WAF composition compared to PAHs and other low solubility UCM components The elevated concentration of oil droplets in media such as CEWAF or HEWAF requires additional steps through chemical analysis and modeling to understand how the dissolved phase composition is changing with the dilutions as diluted droplets dissolve.

span of the test species. Latent or delayed effects can be assessed by including a post-exposure recovery period to evaluate surviving organisms.

2.3.2.1. Duration of exposure. Traditional experimental design and test guidance for toxicity testing of chemicals (Sprague 1969) and oil products have specified a few prescribed observation times (e.g., 24, 48, 72, 96 h) and toxicity endpoints (e.g., 48 or 96 h L(E)C50). In contrast, oil spills are dynamic with exposures that may be more or less than the prescribed reporting times of toxicity tests (e.g., Bejarano et al., 2014). Toxicity studies have demonstrated the reduction of effects concentrations with increasing exposure time for individual compounds (Bailey et al., 1985; Lee et al., 2002)(Mackay et al., 2014) and for oil exposures (Greer et al., 2012; Landrum et al., 2012). More short-duration or pulse test results for individual hydrocarbons would help develop TU-based oil toxicity models that account for the effects of real-world exposure durations. Whole oil bioassays documenting effects of short-duration exposures are needed to validate these models (see French-McCay et al., 2023).

Evaluation of existing data and predictive toxicity model needs has led to a recommendation that acute toxicity tests include the collection of additional time-response data to provide greater utility in estimating adverse effects over time (Sprague 1969; Bejarano et al., 2014). Two options can be used to quantify the observed time-dependence of toxicity. The first involves simply increasing the number of fixed observation periods (e.g., add 2, 4, 8, and 12 hour observations to the conventional 24 hour observation intervals) so that L(E)C50s can be derived at shorter exposure times. Use of experimental designs that include additional observation points earlier in the test allow models to be developed to predict the time course of mortality or other measures of effect (Barron et al., 2008; Sánchez-Bayo 2008). A second approach involves time-to-death or time-to-event (TTD/TTE) experimental designs. In these tests, the time required for each test organism to exhibit the adverse event investigated (e.g., death) is determined. Additionally, the time course of effect (e.g., survival time) can be modelled to provide a continuous estimate of effect over time and extrapolation to longer- and shorter-term exposures (e.g., Barron et al., 2008; Sánchez-Bayo 2008). Time-independent experimental designs such as Incipient Lethal Level (ILL) approaches provide an excellent method to characterize and compare toxic responses without the uncertainty introduced by prescribed observation periods with a defined study termination (Sprague 1969). Methods and applications of modeling time-to-death are provided by Crane et al. (2002) and others (e.g., Sánchez-Bayo 2008; Mackay et al., 2017; French-McCay et al., 2023). The resulting data and model fits obtained using a first-order kinetic model for these two approaches are illustrated in Fig. 2. An advantage of time-to-event test designs is that relevant effects data for modeling can be obtained at high exposure concentrations up to the solubility limit over a short test period.

2.3.2.2. Designs to address time variable and latent effects. In most instances exposures resulting from an oil spill are expected to be acute, and this is commonly mirrored in the duration of the proposed toxicity tests. As mentioned earlier, standard test durations are not representative of the dynamic and ever-changing nature of an oil spill, which cannot only result in shorter exposures than those commonly tested but can also lead to repeated or pulsed exposure scenarios (e.g., related to tidal fluxes) depending on the dynamics of the oil and the behavior of the exposed organisms.

The variable nature of a pulsed exposure, both in terms of magnitude and duration, can lead to different responses than those observed in a continuous exposure of the same duration. Sometimes, a pulsed exposure can result in reduced effects compared to a continuous exposure test, especially if the initial exposure in the pulsed exposure test resulted in the induction of exposure limiting (e.g., mucus production) or detoxification mechanisms (e.g., induction of enzymatic processes). Alternatively, pulsed exposures can also result in increased effects (e.g., if the initial exposure resulted in damage which compromised the ability of the organism to deal with subsequent exposures). Time-to-event designs can also assist in elucidating whether pulsed exposures result in these enhanced or reduced effects by simply comparing the effect level for each exposure scenario at a particular exposure duration (Newman and McCloskey 1996; Reinert et al., 2002).

Recent methodologies address this problem with more mechanistic approaches such as toxicokinetic-toxicodynamic (TKTD) models which assume that a certain effect occurs when an internal threshold concentration occurs. To develop and use these models, information on species specific physiological and life-history traits are needed as well as the parallel estimation of tissue concentrations. This will enable linkage of the observed effects to tissue concentrations and development of models that are able to determine when specific tissue threshold levels might be reached (see Ashauer (2010) for a detailed description of these models, data needs, and advice on experimental design).

Acute and short pulsed exposures can sometimes lead to post-exposure latent or delayed effects. Latent effects are those that are not immediately observable coincident with an environmental exposure, rather some period is required following the initial exposure for observable effects to occur. This delay can be the result of biochemical or physiological processes that result in a cascade of events that ultimately result in a toxic response (e.g., cancer, reproductive effects, target organ toxicity) (e.g., Heintz et al., 2000). One approach used to evaluate the potential for latent effects are studies consisting of an initial short exposure phase, followed by a long-term monitoring phase in clean media. To increase the realism for petroleum spills, the duration and

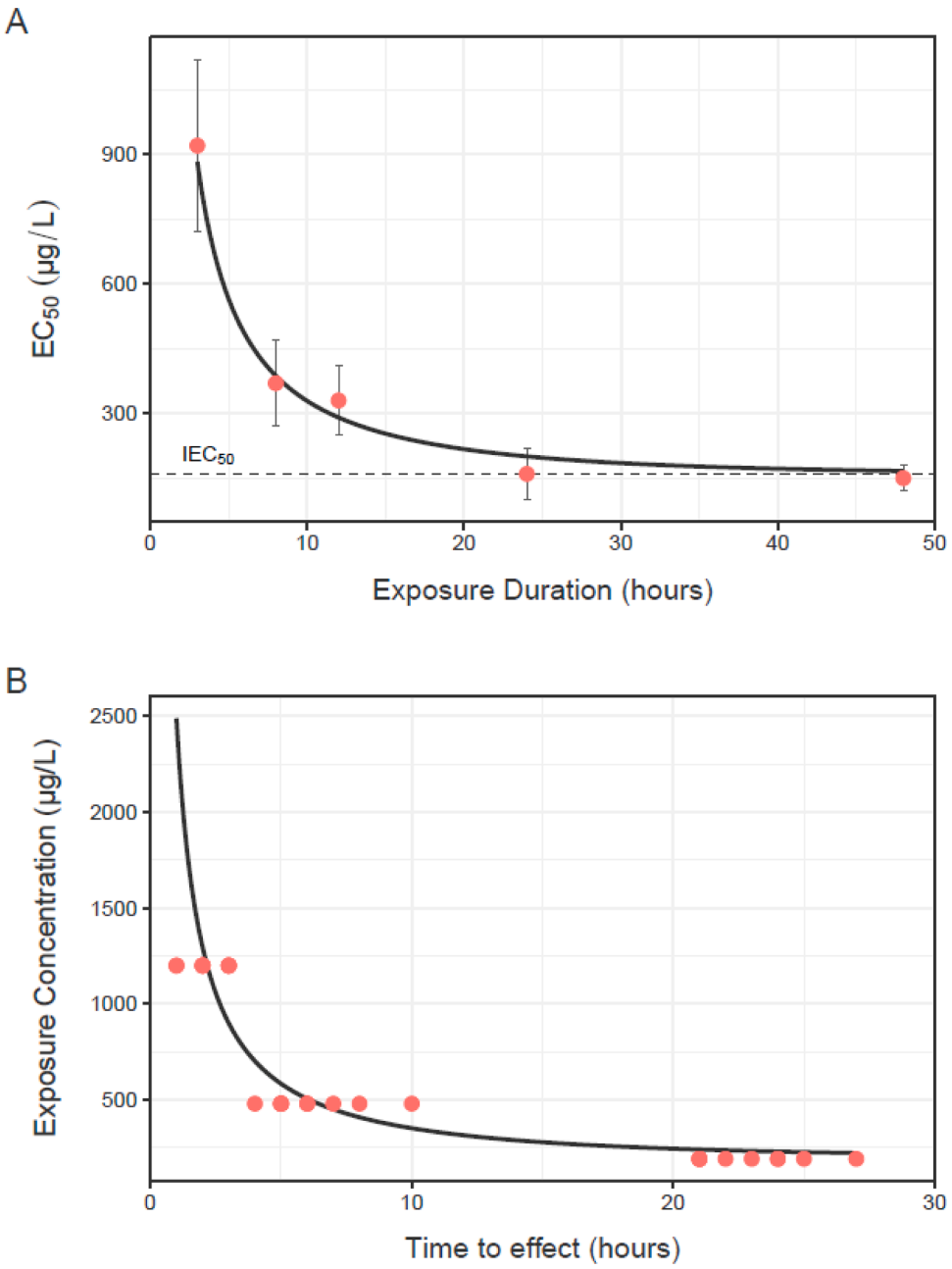


Fig. 2. Examples of time-dependent toxicity data that can be used to support modeling; A) EC50s derived at different exposure durations with time-independent (IEC50) identified as dashed line; and B) Time-to-event for individual test organisms exposed to three exposure concentrations; Solid lines in both graphs denotes model fits assuming first-order kinetics.

intensity of an exposure should be considered in the experimental design. Test exposures should be sufficient to identify and characterize likely adverse effects, but not high enough to be environmentally unrealistic and introduce potential lab-related artifacts (e.g., entrapment of organisms, impediment to gas exchange etc.). Experimental durations should be sufficient to characterize appropriate acute or chronic exposures depending on field-observed conditions. Tests to address latent or delayed effects can also consider multi-generational studies where only the initial parent generation is exposed, but subsequent generations are monitored for toxic effects.

For endpoints such as population growth, often used in tests with primary producers (e.g., algae), it is common to observe reduction in population size after exposure due to the death of a proportion of individuals. However, this reduction in population may be temporary if the exposure did not lead to long-lasting effects in the surviving organisms. In this case, exposure-recovery studies would be recommended that include a short exposure phase, followed by a monitoring phase. Similar to the approach described above, but where the monitoring phase would assess the capacity of the exposed organisms to return to some pre-defined "ecologically recovered" condition (Gergs et al., 2016).

3. Guidance for laboratory toxicity testing

Experimentally derived toxicity data are extremely valuable in providing key qualitative and quantitative information for spill response planning such as Net Environmental Benefits Analysis (NEBA)/Spill Impact Mitigation Assessment (SIMA) (IPIECA-IOGP, 2015) and hazard assessment. It allows for the evaluation of relative hazards from different SRAs and oils, determination of species or taxa sensitivity to the product,

assessment of the effects of oil with/without SRA and supports the screening or regulatory approval of response agents (Bejarano and Mearns 2015; Redman and Parkerton 2015). There is no single or universal experimental design for a toxicity test that will adequately address all objectives. However, whether the test is geared towards hazard classification (Fig. 3A), hypothesis testing (Fig. 3B), or informing biological effect models (Fig. 3C), there are common threads which run through all oil testing.

Here, some general guidance and considerations are provided for the conduct of aquatic toxicity tests using oil and oil derived products, which are broadly applicable regardless of experimental objective.

3.1. Species selection

In an ideal world, toxicity data would be available for all chemicals and chemical mixtures, and all species and life stages resident in an area, thus making it possible to assess potential risks resulting from chemical exposure with minimal uncertainty. Unfortunately, data are usually available for only a fraction of the species comprising an ecosystem; frequently this requires us to extrapolate data from common laboratory "surrogate" species to site-specific resident species. To appropriately understand the impacts to aquatic ecosystems, it is recommended to consider species representing different trophic levels and feeding types (filter, deposit, detritivore, etc.), thus encompassing different exposure routes. Life-stage selection is also an important consideration. Many advantages have been noted for testing with early life-stages of longer-lived organisms due to their potential increased sensitivity attributed to greater chemical adsorption/uptake or reduced toxifying metabolic capabilities, shorter test duration, and smaller organism size making it possible to test under laboratory conditions (i.e., smaller test containers and less volume of generated waste).

As described, test species selection is an important consideration in oil toxicity testing and should be related to the specific purpose of the testing. An ideal toxicity test species might meet one or more of the following criteria: (1) sensitivity, (2) widely distributed, abundant, (3) indigenous or representative, (4) recreationally, commercially and/or ecologically important, (5) laboratory tolerant, easy to culture, and (6) life history. Often there is a desire to test site-specific resident species; however, these organisms are likely to have to be field-collected and brought into the laboratory. There are pros and cons to the use of field-collected vs. laboratory-cultured animals. Field-collected animals can be relevant to local spill issues and contingency planning, and responses to test chemicals may be more realistic due site-specific issues. There are, however, considerations including restrictions on collection (e.g., required permits), holding time, health (e.g., seasonal differences), etc., and unknown history of exposure to toxic substances. Cultured animals have the advantage of being readily available, genetically consistent, a known exposure history, and known health status. Standardized toxicity tests have been developed by many organizations (e.g., OECD, ASTM, USEPA, ECCC), and these guidelines generally contain requirements for the care, culture, and testing of recommended test species. Criteria for culture health prior to testing and validity criteria for lab controls and water quality at the end of the test have been established, and these factors when combined with other QA/QC practices (such as reference toxicant testing) provide confidence in test results. Standardized test guidelines with defined test species require reviewing culture colony health in the 7–14 days before testing and delaying the test if the colony appears stressed or mortality exceeds method specifications. Standard guidelines provide best practices for culture colony health, control mortality and QA/QC (Weber 1991). For non-standard organisms, including field collected organisms, aquaculturists may be able to provide useful information regarding natural die-offs during development so organism health can be appropriately assessed prior to test initiation.

Regardless of whether they are field-collected or cultured, the test organisms should all be from the same source, of a uniform size, age, and physiological condition, and devoid of visible disease and parasites. Care should be taken to minimize collection and handling stress, and organisms should be gradually acclimated to test conditions. Acclimation time will vary by species and standard testing protocols. Typical holding acclimation times for field collected organisms range from 7 to 10 days prior to testing; however, this is typically only applicable to short-term acute tests with older organisms; tests that begin with embryonic or larval organisms cannot meet this criterion due to their rapid development.

In evaluating spill impacts, representative species can be surrogates for local species or species at risk, which are easier to source, care for, and/or evaluate effects. They may not, however, include all the characteristics of the target species. Where comprehensive species data are not available, using species traits to estimate impact may be possible and appropriate. In some cases, specific traits for species or life stages may

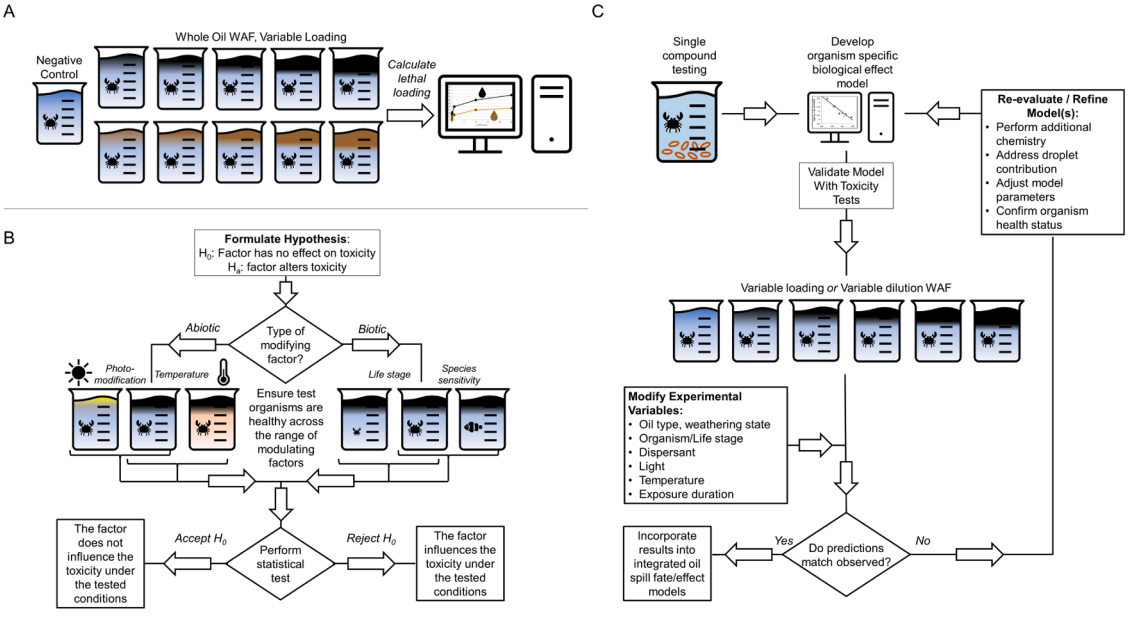


Fig. 3. Examples of experimental designs based on differing objectives. A) Substance hazard classification, B) hypothesis testing, and C) informing biological effect models.

need to be considered when evaluating risks. For example, as noted in Alloy et al. (2023), assessment of photosensitization may require use of translucent early life stages of aquatic organisms because of their increased sensitivity.

Species with commercial value bridge the socioeconomic and environmental considerations; local stakeholder groups will be particularly interested in their protection and recovery in the event of a spill. In some cases, threatened or endangered species are of particular interest. Special considerations are needed for testing protected species and alternate techniques, e.g., testing with surrogate species, must be considered since they may provide the only means to obtain data. Also, threatened/endangered species may be considered differently in a NEBA because impacts at the individual rather than the population level are considered important. Comparison of species sensitivity for a range of contaminants has shown that species listed as threatened and endangered are generally no more sensitive than non-listed species (Raimondo et al., 2008). The taxonomic grouping of the organism, organism traits, and chemical mode of action appear to be most important in determining intrinsic species sensitivity, rather than geographic distribution or rarity of the species (Rico and Van den Brink, 2015). Toxicity data available for commonly tested species can also be used as input to interspecies correlation (ICE) models to obtain toxicity predictions for untested species so that these estimates can be used to increase available data or toxicity estimates for substances with limited hazard data (Bejarano and Barron 2014; Bejarano 2019). If hazard data are restricted to a fixed exposure duration, SSDs and corresponding HC5 (hazardous concentration for 5% of species) values can be derived for different exposure durations (Bejarano 2018).

Finally, many jurisdictions are restricting the use of toxicity testing with vertebrates for the purpose of chemical hazard assessments. Testing with invertebrates, molecular tools, cell lines, unprotected life stages of vertebrates or other new approach methodologies may be used to replace vertebrate testing (e.g., Yu et al., 2005). However, when modeling effects for contingency planning or damage assessments, having toxicity data available for the fish species in question adds confidence and relevance to the models.

3.2. Inclusion of test controls

As with all toxicity testing, it is imperative that adequate controls be included in oil spill toxicity assessments. A negative control (i.e., no-oil media control) is required to ensure healthy baseline response of test animals. For CEWAF tests, the use of a dispersant control at the highest dispersant concentration used to prepare the CEWAF may be employed. It is also recommended that a positive control be included. A positive control can be useful to control for variability in responses between tests and batches of organisms (Olsen et al., 2011), and the results from side-by-side positive control tests build a database of known sensitivity thresholds and reproducibility that can be used to establish coefficients of variation for an organism. For acute oil toxicity testing, we recommended that 1-methylnaphthalene be used as a positive control, in addition to any species specific standard reference toxicant (e.g., KCl), or method specific control (e.g., solvent control). 1-methylnaphthalene offers the advantages of being commercially available, sufficiently toxic to define a exposure:response relationship, largely present in oil, easy to work with (e.g., a water- soluble liquid with a strong fluorometric signal), and there is a growing database of species sensitivity to this compound. Generating an EC/LC50 for 1-methylnapthtalene, and then including that concentration as a positive control for the species of interest in further testing (either with other individual compounds or whole oils), will provide an appropriate measure of quality assurance and control and can potentially replace a multi-concentration reference toxicant testing program. Adoption of this standardized approach would allow comparable data to be generated that can be used within spill response and risk assessment modeling.

3.3. Exposure regime

Exposure system designs may be static, static-renewal, intermittent flow-through, continuous recirculating, or continuous flow-through. Each has advantages and disadvantages, and selection of system design is guided by the solubility, volatility and degradability of hydrocarbons and the goals of the test (Parkerton et al., 2023a). The overall stability of the test chemical(s) is a consequence of the exposure system and is influenced by a number of physical, chemical, and biological factors. Test systems should minimize chemical loss as a result of adsorption to exposure system surfaces; this can be mitigated by the use of low-adsorption components and materials and by preconditioning/equilibration of the test system for a period of time determined by the results of stability tests (OECD 2019a). Covered exposure vessels constructed of non-reactive materials are recommended, coupled with methods that limit loss of volatile components and that support temporally stable and analytically verifiable exposure concentrations. Losses may also occur due to partitioning to suspended particulates (food particles or detritus) and to the test organisms themselves; thus, cleaning and removal of detritus, as appropriate, or testing using methods that mitigate these concerns, e.g., flow-through systems should be employed.

Minimizing concerns with oxygen depletion due to BOD/COD loss can be addressed through aeration or flow-through test designs. Oxygen depletion by organisms and ammonia waste concerns can be addressed by following recommended biological loading guidance from the various standard methods.

3.4. Exposure characterization

Characterizing the composition and concentration of the individual hydrocarbons in exposure solutions is critical to the interpretation of toxicity data and the evaluation of environmental hazards posed by chemicals. Petroleum and petroleum-derived products pose a unique problem for toxicity testing because the presence and stability of the hydrocarbon components is dictated by their individual chemical characteristics. This, in turn, leads to temporal changes in the composition and toxicity of hydrocarbon solutions in the laboratory and in spill incidents. If the experimental design involves exposures that decline (i.e., static, static-renewal, pulsed) details of the decline must be quantified and the frequency of renewals reported. Acceptable ways to quantify declining exposures (e.g., assume exponential decay curve and quantify at the initial time and at the end of the exposure or prior to a renewal) are described and guidance provided for calculating the concentration used for effect assessment based on exposure approach (OECD 2019a). Regardless of exposure system, chemical determinations of the exposure solution components or marker compounds of the highest and lowest test concentration (or lowest quantifiable concentration) and component concentrations around the expected E/LC50 are considered a minimum requirement. Recommendations on the frequencies of analytical measurements are provided in OECD (2019a) and these should be followed in combination with the media specific properties. Some general recommendations are illustrated in Fig. 4 for three different exposure regimes.

Monitoring exposures is key to producing useful data for validating models, comparative hazard assessment and model validation, with consideration given to losses during sampling and sample storage before analysis. Inadequate characterization of exposure concentrations prohibits the use of the data for comparative or modeling purposes, and represents a significant waste of resources, time, and animals (see Dettman et al., 2023and Bejarano et al., 2023). However, no single analytical technique can provide a complete determination of all the hydrocarbons present in a water sample, so steps must be taken to plan for the proper analyses and to take and preserve samples for analysis. Analytical techniques to quantify different hydrocarbons dictate sampling intervals and this must be considered in designing experimental

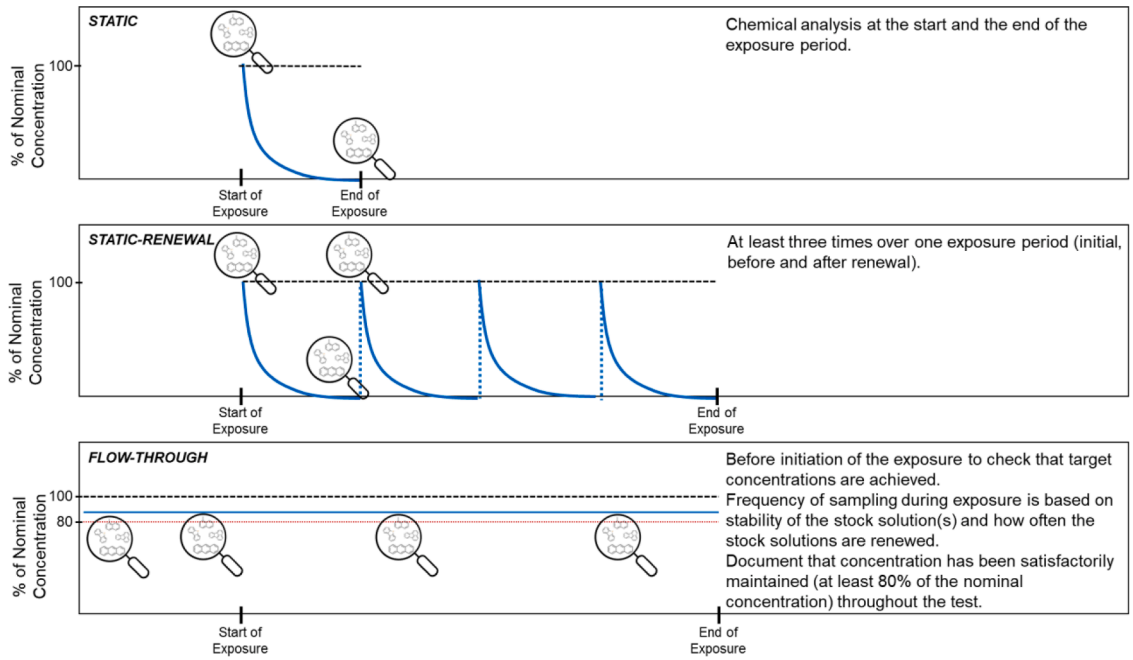


Fig. 4. Overview of suggested analytical characterization frequency based on different exposure regimes.

approaches. Guidance on frequency of analytical determinations, measurements, and data handling are provided in OECD guidelines (e.g., (OECD 2019b, 2019a, 1992)), and is dependent on the media, exposure regime, and experimental objective. Methods for analysis, including minimum requirements, and considerations for sampling, sample storage, and handling are discussed in Dettman et al. (2023).

Spill response agents present a unique problem for the analysis and interpretation of toxicity data. SRAs are complex mixtures of surfactants (many of which are proprietary) and hydrocarbon solvents, there are no simple and consistent means for measuring dispersant or dispersant components in water (Fingas, 2017). Generally, efforts have focused on the quantification of 1–2 active ingredients reported in safety data sheets as target analytes. However, caution should be taken in making generalizations about the potential for only one or a few compound(s) to serve as useful markers or indicators of exposure (CRRC, 2012). A discussion of the challenges and research need associated with SRA characterization is presented in Dettman et al. (2023)

3.5. Toxicity endpoints

Endpoints (e.g., development, survival, growth, reproduction) are used in dose-response modeling to estimate effect thresholds including NOEC, LOEC, MATC (No observed effect concentration, Lowest observed effect concentration, Maximum acceptable toxicant concentration) and ECx/LCx (Effect concentration or Lethal concentration at a defined percent (x)). For ECx determination, the study endpoint must be clearly defined and relevant to the test species. Selection of endpoints in an aquatic toxicity study should consider the exposure regime (e.g., acute vs chronic study) and species and life stage. Typically, acute tests have only measured organism immobilization or mortality, whereas longer duration chronic tests measure organism growth and/or reproduction.

Acute and chronic toxicity data are often used to generate SSDs, which allow the comparison of the relative sensitivity of Multiple species to the same chemical or mixture. Effect thresholds, e.g., HC5, derived from SSDs represent scientifically defensible benchmarks, as verification of consensus values for levels of concern (Bejarano and Mearns 2015). Inclusion of data for an SSD requires the study and endpoints are consistent and adequately reported (Bejarano et al., 2023), and properly characterized (Dettman et al., 2023).

Measurement endpoints that can be directly related to population-level effects tend to be considered the most relevant and include survival, growth, or reproduction. Other potential endpoints can be related to sublethal observations that can span a wide range of biological responses, from changes in behavior to cellular, subcellular, or biochemical changes that may be related to environmental exposures. Sublethal endpoints may be important and may be considered, but these approaches must be shown to have relevance to “traditional” endpoints and population-level concerns (Doering et al., 2019). The Adverse Outcome Pathway (AOP) analysis approach may provide a means to interpret these observations in the future (Ankley et al. 2010); however, clear relationships must be established between the observed effect and the cascade of steps that lead to population impacts.

3.6. Exposure metrics

When reporting toxicity endpoints, the choice of exposure metric can have implications for the interpretation of the results. This is discussed in greater detail in Parkerton et al. (2023b), and briefly mentioned here for completeness. Expressing toxicity as%WAF or%CEWAF has significant limitations since it has a limited relationship to the hydrocarbon composition of the solution and should be avoided in most experimental applications. Summary measures like total PAH (ΣPAH) and Total Petroleum Hydrocarbons (TPH) have previously been used but again they are not specific to the mixture and significant differences in toxicity have been observed for solutions which have the same ΣPAH or TPH values. This is because not all hydrocarbons are equally toxic; therefore, toxic potency of the individual components should ideally be considered in any exposure metric to permit comparison between hydrocarbon mixtures and to assess solution toxicity. Reporting concentrations of the various fractions of the TPH by distillation cut and on a percent aliphatic and aromatic basis has been recommended as an approach for whole oils (Dettman et al., 2023). Normalizing the concentration of individual hydrocarbon components in a solution based on their relative toxicity (Toxic Units (TU), Hermans and Leeuwangh, 1982, Parkerton et al., 2023b) permits the summation of concentrations while considering the toxic contribution of the components. This is the approach used by most current predictive toxicity models; further detail is provided in

French-McCay et al. (2023).

3.7. Statistical analysis of toxicity data

Appropriate statistical analysis of toxicity data informs experimental design and conduct, and adherence to recommended analytical methods ensures correct interpretation of results. Similarly, appropriate design can improve both the accuracy and precision of test data and subsequently determined threshold values. Adequate randomization and avoidance of pseudo replication is required to meet the assumptions of parametric tests. Negative, positive and solvent controls where appropriate, e.g., in single hydrocarbon tests, must be included in the design. Where multiple controls are used, several analytical approaches may be used to evaluate potential solvent effects. Recommendations for the most appropriate control or combination of control groups to compare with treatment groups vary amongst regulatory agencies (see OECD, 2006 and Green et al., 2018 for an evaluation of different approaches). Statistically derived threshold values are determined from a variety of endpoint types (e.g., quantal, continuous, or discrete), and different analyses are appropriate for each. OECD (2006) and Environment Canada (2007) include flowcharts and provide guidance on the selection of statistical approaches, and describes assumptions and limitations for hypothesis testing, concentration-response modeling, and biology-based methods.

Determination of environmentally protective concentrations for environmental contaminants (e.g., Predicted No Effect Concentration (PNEC), Final Chronic Value (FCV)) involve consideration of the range of sensitivity to a contaminant across M. species representing a range of trophic levels and life strategies. Typically, these values are derived from single-species laboratory ecotoxicological tests measuring specific endpoints (typically survival, reproduction, or growth). The SSD approach is used to provide community-level protection. Calculation of appropriate SSD endpoints is beyond the scope of this discussion; however, discussion of the approach and methods for calculation of these values can be found in ECHA (2008), Stephan et al. (1985), EFSA (2015), and Fox et al. (2021).

4. Summary and recommendations

The approach to risk assessment has shifted from attempting to reproduce field conditions and exposures, to the application and development of consistent test methods needed for calibration and

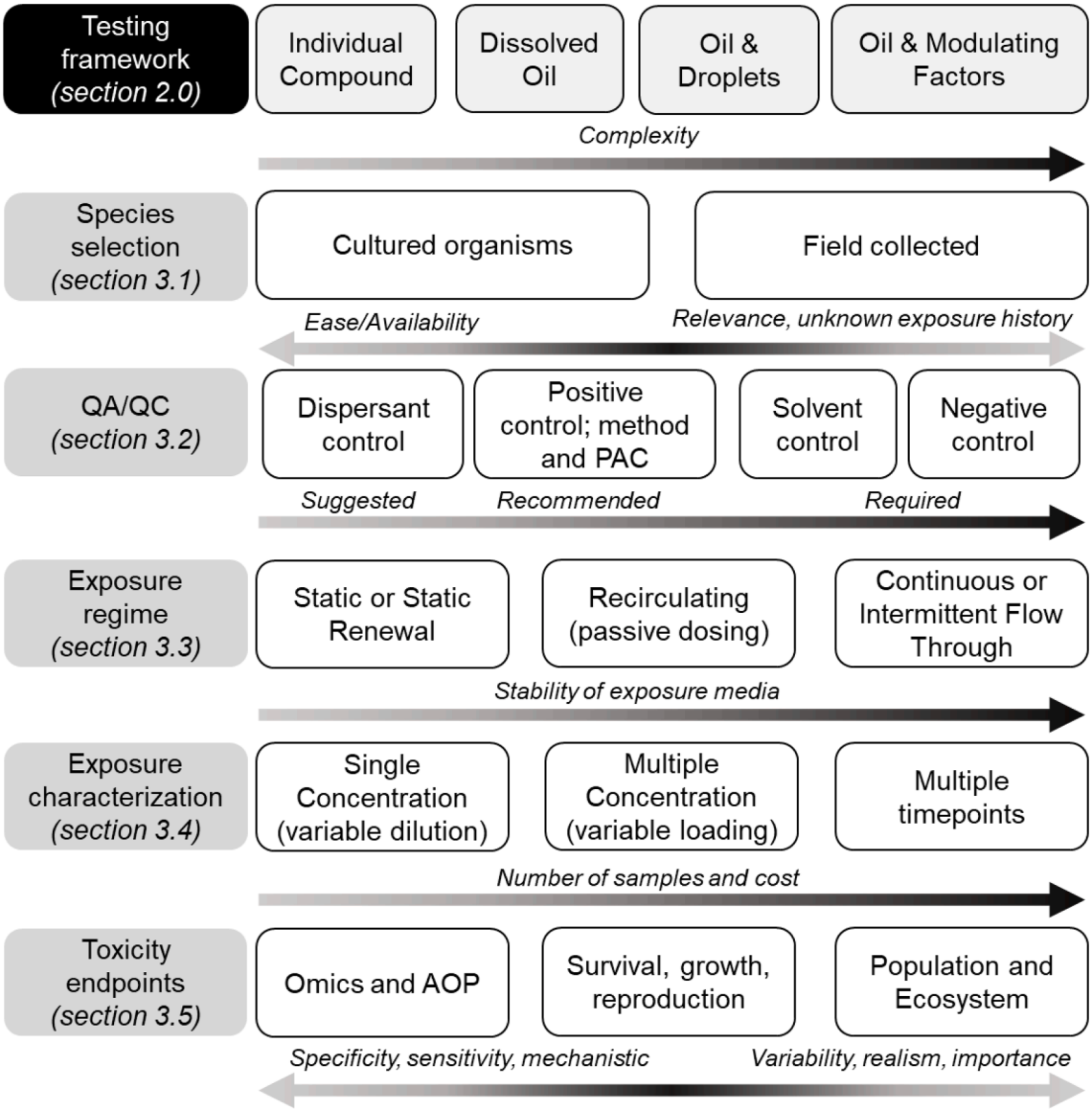


Fig. 5. Summary of key issues for consideration in the design and conduct of oil toxicity experiments.

validation of toxicity models. This requires standardized testing approaches that provide reproducible test results and comparisons of product hazards which can be used to evaluate the risks of environmentally realistic exposures (Hodson et al., 2019; National Academies of Sciences and Medicine 2020). As greater confidence is gained in the reliability of such models for predicting effects in lab toxicity studies, these models may be more credibly applied for effect assessment under a wide range of oil exposure scenarios obtained using either field measurements (if such sampling is feasible) or spill fate model predictions.

An overview of the some of the key issues discussed in this paper are presented in Fig. 5; oil toxicity practitioners are strongly encouraged to consult this figure and consider adopting these recommendations when designing experiments.

The relevance of this approach relies on the application of standardized test designs and methodology to ensure comparability of results across studies. Guidelines for standardized testing also allow the assessment of the relevance of studies conducted with non-standard species or exposures.

Following the principles and guidance outlined in this paper can greatly improve the utility of laboratory-based toxicity testing for use in hazard assessment and for the validation of oil toxicity models. Experimental validation and improvement of model predictions represents a significant step forward in the ability to respond to and mitigate the biological effects of oil spills in the environment.

4.1. Research needs

A number of continuing research needs have been identified:

- Additional high-quality, single-species acute and chronic toxicity data are needed for an array of parent and substituted PAHs, as well as other aliphatic, aromatic and polar compounds. These tests should follow standardized test methods with appropriate analytical characterization of exposures. These data will provide the necessary support for toxicity predictions currently based on chemical characteristics (e.g., octanol:water portioning), assumptions regarding mode of toxic action (e.g., narcosis), and the relationship between acute and chronic toxic responses. This information will ultimately reduce the uncertainty associated with current predictive toxicity models (e.g., McGrath et al., 2018).
- Petroleum spills frequently lead to intermittent hydrocarbon exposures that can vary dramatically in composition, duration, and magnitude. Targeted data are needed that allow further characterization of the time-course of toxicity across species/endpoints and hydrocarbons/SRA so that effect concentrations can be calculated and validated for the wide-range of time-varying exposure durations relevant to oil spills (French McCay et al. 2023; Parkerton et al., 2023b).
- Toxicity testing of SRAs and SRA/hydrocarbon mixtures present specific methodological difficulties that must be addressed through standardization of the SRA testing methods. Furthermore, regulations should require identification and characterization of the toxic properties of “proprietary” components in SRA mixtures and analytical methods for measuring SRA components must be provided.
- In all studies, bioassay conditions need to be well-characterized (i.e., with analytically verified exposure concentrations over the duration of the experiment, Dettman et al., 2023) for the data to be useful for modeling, impact assessment, comparability, and interpretation of results. The species and life stages tested should include both sensitive and insensitive ones, as exposure duration and effects of modifying factors may vary.
- Development of models that can address sublethal exposures and latent effects (i.e., “omic” endpoints; (Brodersen 1987; Zhao and Newman 2004)). Clear relationships must be established between the observed effect and the cascade of effects that lead to population impacts. Single-analyte bioassays would elucidate mechanisms accounting for delayed effects.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

No data was used for the research described in the article.

Acknowledgements

This work was supported through the Government of Canada’s Department of Fisheries and Oceans Multi-Partner Research Initiative. The manuscript has been subjected to U.S. Environmental Protection Agency review and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

References

Adams, J., Charbonneau, K., Tuori, D., Brown, R.S., Hodson, P.V., 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. Can. Sci. Advis. Secret.

Alloy, M., Finch, B., Ward, C., Redman, A., Bejarano, A.C., Barron, M., 2023. Recommendations for Advancing Test Protocols Examining the Photo-induced Toxicity of Petroleum and Polycyclic Aromatic Compounds. Aquat. Toxicol. 256, 106390 https://doi.org/10.1016/j.aquatox.2022.106390.

Anderson, J.W., Neff, J.M., Cox, B.A., Tatem, H.E., Hightower, G.M., 1974. Characteristics of Dispersions and Water-Soluble Extracts of Crude and Refined Oils and Their Toxicity to Estuarine Crustaceans and Fish 88, 75–88.

Ankley, G.T., et al., 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ. Toxicol. Chem. 29, 730–741.

Ashauer, R., 2010. Toxicokinetic–toxicodynamic modelling in an individual based context—Consequences of parameter variability. Ecol. Modell. 221, 1325–1328. https://doi.org/10.1016/j.ecolmodel.2010.01.015.

Aurand, D., Coelho, G., 2005. Cooperative Aquatic Toxicity Testing of Dispersed Oil and the “Chemical Response to Oil Spills: Ecological Effects Research Forum (CROSERF)” A Model for Cooperative Research by Industry and Government ’Ecosystem Management & Associates, Inc. Technical Report 07-03. Ecosystem Management & Associates, Inc., Lusby, MD.

Bailey, H.C., Javitz, H.A., Liu, D.H.W., 1985. Time/Toxicity Relationships in Short-Term Static, Dynamic, and Plug-Flow Bioassays/. ASTM International, West Conshohocken, Pa. https://doi.org/10.1520/STP33573S.

Barron, M.G., Raimondo, S., Russom, C., Vivian, D.N., Yee, S.H., 2008. Accuracy of chronic aquatic toxicity estimates determined from acute toxicity data and two time-response models. Environ. Toxicol. Chem. 27, 2196–2205. https://doi.org/10.1897/08-004.1.

Barron, M.G., Hemmer, M.J., Jackson, C.R., 2013. Development of Aquatic Toxicity Benchmarks for Oil Products Using Species Sensitivity Distributions 9, 610–615. 10.1002/ieam.1420.

Bejarano, A.C., Barron, M.G., 2014. Development and practical application of petroleum and dispersant interspecies correlation models for aquatic species. Environ. Sci. Technol. 48, 4564–4572. https://doi.org/10.1021/es500649v.

Bejarano, A.C., Mearns, A.J., 2015. Improving environmental assessments by integrating Species Sensitivity Distributions into environmental modeling : examples with two hypothetical oil spills. Mar. Pollut. Bull. 93, 172–182. https://doi.org/10.1016/j.marpolbul.2015.01.022.

Bejarano, A.C., Clark, J.R., Coelho, G.M., 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. https://doi.org/10.1002/etc.2501.

Bejarano, A.C., Adams, J.E., McDowell, J., Parkerton, T.F., Hanson, M.L., 2023. Recommendations for improving the reporting and communication of aquatic toxicity studies for oil spill planning, response, and environmental assessment. Aquat. Toxicol. 255 https://doi.org/10.1016/j.aquatox.2022.106391.

Bejarano, A.C., 2018. Critical Review and Analysis of Aquatic Toxicity Data on Oil Spill Dispersants. Environ. Toxicol. Chem. 37, 2989–3001. https://doi.org/10.1002/etc.4254.

Bejarano, A.C., 2019. Further Development and Refinement of Interspecies Correlation Estimation Models for Current-Use Dispersants. Environ. Toxicol. Chem. 38, 1682–1691. https://doi.org/10.1002/etc.4452.

Black, T.A., Hanson, M.L., Palace, V.P., Rodriguez-Gil, J.L., 2021. Surface-Dwelling Aquatic Insects in Low-Energy Freshwater Environments Are Highly Impacted by Oil Spills and the Surface Washing Agent Corexit EC9580A Used in Oil Spill Response. Environ. Toxicol. Chem. 40, 1298–1307. https://doi.org/10.1002/etc.4976.

Bobra, A.M., Shiu, W.Y., MacKay, D., 1983. Toxicity of Hydrocarbons and Crude Oils, Environment Canada, Ottawa, ON, pp. 52-58, 1983. In 'Proceedings of the Sixth AMOP Technical Seminar'. Ottawa, ON, pp. 52–58.

Brodersen, C., 1987. Rapid narcosis and delayed mortality in larvae of king crabs and kelp shrimp exposed to the water-soluble fraction of crude oil. Mar. Environ. Res. 22, 233–239. https://doi.org/10.1016/0141-1136(87)90038-9.

Buikema, A.L., Niederlehner, B.R., Cairns, J., 1982. Biological monitoring part IV—Toxicity testing. Water Res. 16, 239–262. https://doi.org/10.1016/0043-1354(82)90188-9.

Canada, Environment, 2007. Guidance Docunment on Statistical Methods for Environmental Toxicity tests. Report EPS 1/RM/46. Method development and Applications Section. Environment Canda, Ottawa, Canada, p. 241.

Crane, M., Newman, M.C., Chapman, P., Fenlon, J., 2002. Risk Assessment with Time-to-Event Models. CRC Press, Boca Raton, FL, USA.

CRRC, RPI, NOAA. 2012. The Future of Dispersant Use in Oil Spill Response Initiative. https://crrc.unh.edu/sites/default/files/migrated_unmanaged_files/dispersant_future_11/Dispersant_Initiative_FINALREPORT.pdf.

Dettman, H., Wade, T.L., French-McCay, D.P., Bejarano, A.C., Hollebone, B.P., Faksness, L., Mirnaghi, F.S., Yang, Z., Loughery, J.R., Pretorius, T., de Jourdan, B., 2023. Recommendations for the advancement of oil-in-water media and source oil characterization in aquatic toxicity studies. Aquat. Toxicol. (Submitted).

Di Toro, D.M., Mcgrath, J.A., 2000. Technical basis for narcotic chemicals and polycyclic aromatic hydrocarbon criteria. II. Mixtures and sediments. Environ. Toxicol. Chem. 19, 1971–1982. https://doi.org/10.1002/etc.5620190804.

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. Toxicol. Chem. 26, 24–36. https://doi.org/10.1897/06174r.1.

Doering, J., Hecker, M., Villeneuve, D., Zhang, X., 2019. Adverse Outcome Pathway on aryl hydrocarbon receptor activation leading to early life stage mortality, via increased COX-2. 10.1787/bd46b538-en.

Echols, B.S., 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. https://doi.org/10.1007/s00244-015-0228-6.

EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2015. Scientific Opinion on the effect assessment for pesticides on sediment organisms in edge-of-field surface water. EFSA J. 13 (7), 4176. https://doi.org/10.2903/j.efsa.2015.4176, 2015145 pp.

European Chemicals Agency (ECHA), 2008. Guidance on Information Requirements and Chemical Safety Assessment Chapter R.10: Characterisation of Dose [concentration]-response for Environment. European Chemicals Agency, Helsinki, Finland, p. 65.

Fieldhouse, B.G., Jackman, P., Alijawahari, M., Bauer, M., Hopper, C., Raats, A., Singh, N., 2019. An Alternative Approach to Aquatic Toxicity Assessment of Dispersants to Evaluate Suitability for Regulatory Listing in Canada as an Oil Spill Treating Agent. In 'Proceedings of the Forty-Second AMOP Technical Seminar'. Environment and Climate Change Canada, Ottawa, ON, pp. 610–625.

Fingas, M., 2017. A review of the literature related to oil spill dispersants, p. 164. https://www.pwsrcac.org/wp-content/uploads/filebase/programs/environmental_monitoring/dispersants/A%20Review%20of%20Literature%20Related%20to%20Oil%20Spill%20Dispersants,%20September%202017.pdf.

Forth, H.P., Mitchelmore, C.L., Morris, J.M., Lipton, J., 2017. 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. https://doi.org/10.1002/etc.3672.

Foster, K.L., Mackay, D., Parkerton, T.F., Webster, E., Milford, L., 2005. Five-stage environmental exposure assessment strategy for mixtures: gasoline as a case study. Environ. Sci. Technol. 39, 2711–2718. https://doi.org/10.1021/es048734p.

Fox, D.R., van Dam, R.A., Fisher, R., Batley, G.E., Tillmanns, A.R., Thorley, J., Schwarz, C.J., Spry, D.J., McTavish, K., 2021. Recent developments in species sensitivity distribution modeling. Environ. Toxicol. Chem. 40 (2), 293–308.

French-McCay, D., Crowley, D., Rowe, J.J., Bock, M., Robinson, H., Wenning, R., Hayward, A., Joeckel, J., Nedwed, T.J., Parkerton, T.F., 2018. Comparative Risk Assessment of spill response options for a deepwater oil well blowout: part 1 . Oil spill modeling. Mar. Pollut. Bull. 0–1. https://doi.org/10.1016/j.marpolbul.2018.05.042.

French-McCay, D.P., Jayko, K., Li, Z., Spaulding, M.L., Crowley, D., Mendelsohn, D., Horn, M., Isaji, T., Kim, Y.H., Fontenault, J., Rowe, J.J., 2021. Oil fate and mass balance for the Deepwater Horizon oil spill. Mar. Pollut. Bull. 171, 112681 https://doi.org/10.1016/j.marpolbul.2021.112681.

French-McCay, D.P., Parkerton, T.F., de Jourdan, B., 2023. Bridging the Lab to Field Divide: advancing Oil Spill Biological Effects Models Requires Revisiting Aquatic Toxicity Testing. Aquat. Toxicol. 256, 106389. https://doi.org/10.1016/j.aquatox.2022.106389.

French-McCay, D.P., 2002. Development and application of an oil toxicity and exposure model, OilToxEx. Environ. Toxicol. Chem. 21, 2080–2094. https://doi.org/10.1002/etc.5620211011.

Gergs, A., Classen, S., Strauss, T., Ottermanns, R., Brock, T.C.M., Ratte, H.T., Hommen, U., Preuss, T.G., 2016. Ecological Recovery Potential of Freshwater Organisms: Consequences for Environmental Risk Assessment of Chemicals BT - Reviews of Environmental Contamination and Toxicology Volume 236 (Ed. De Voogt, P.). Springer International Publishing, Cham, pp. 259–294. https://doi.org/10.1007/978-3-319-20013-2_5.

Green, J.W., Springer, T.A., Holbech, H., 2018. Statistical Analysis of Ecotoxicity Studies. John Wiley & Sons.

Greer, C.D., Hodson, P.V., Li, Z., King, T., Lee, K., 2012. Toxicity of crude oil chemically dispersed in a wave tank to embryos of Atlantic herring (Clupea harengus). Environ. Toxicol. Chem. 31, 1324–1333. https://doi.org/10.1002/etc.1828.

Heintz, R.A., Rice, S.D., Wertheimer, A.C., Bradshaw, R.F., Thrower, F.P., Joyce, J.E., 2000. Delayed effects on growth and marine survival of pink salmon Oncorhynchus gorbuscha after exposure to crude oil during embryonic development. Mar. Ecol. Prog. Ser. 208, 205–216.

Hermans, J., Leeuwangh, P., 1982. Joint Toxicity of Mixtures of 8 and 24 Chemicals to the Guppy (*Poecilia reticulata*). Ecotoxicol. Environ. Safety 6, 302–310.

Hodson, P.V., Adams, J., Brown, R.S., 2019. Oil toxicity test methods must be improved. Environ. Toxicol. Chem. 38, 302–311. https://doi.org/10.1002/etc.4303.

IPIECA-IOGP, 2015. Response strategy development using net environmental benefit analysis (NEBA). Good pratice guidelines for incident management and emergency response personnel. IOGP Rep. 527.

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 hydrocarbon and petroleum hydrocarbon case studies. Integr. Environ. Assess. Manag. 8, 217–230. https://doi.org/10.1002/ieam.277.

Langdon, B.S.E.C.J., Rand, W.A.S.G.M., 2018. A Comparative Assessment of the Aquatic Toxicity of Corexit 9500 to Marine Organisms. Arch. Environ. Contam. Toxicol. https://doi.org/10.1007/s00244-018-0568-0.

Lee, J.-H., Landrum, P.F., Koh, C.-H., 2002. Toxicokinetics and Time-Dependent PAH Toxicity in the Amphipod Hyalella azteca. Environ. Sci. Technol. 36, 3124–3130. https://doi.org/10.1021/es011201l.

Mackay, D., McCarty, L.S., Arnot, J.A., 2014. Relationships between exposure and dose in aquatic toxicity tests for organic chemicals. Environ. Toxicol. Chem. 33, 2038–2046. https://doi.org/10.1002/etc.2649.

Mackay, D., Celsie, A.K.D., Parnis, J.M., McCarty, L.S., Arnot, J.A., Powell, D.E., 2017. The chemical exposure toxicity space (CETS) model: displaying exposure time, aqueous and organic concentration, activity, and onset of toxicity. Environ. Toxicol. Chem. 36, 1389–1396. https://doi.org/10.1002/etc.3668.

McGrath, J.A., Fanelli, C.J., Di Toro, D.M., Parkerton, T.F., Redman, A.D., Paumen, M.L., Comber, M., Eadsforth, C.V., den Haan, K., 2018. Re-evaluation of target lipid model–derived HC5 predictions for hydrocarbons. Environ. Toxicol. Chem. 37, 1579–1593. https://doi.org/10.1002/etc.4100.

National Academies of Sciences and Medicine, E., 2020. The Use of Dispersants in Marine Oil Spill Response 'The Use of Dispersants in Marine Oil Spill Response. The National Academies Press, Washington, DC. https://doi.org/10.17226/25161.

National Academies of Sciences, Engineering, and Medicine, 2020. The Use of Dispersants in Marine Oil Spill Response. The National Academies Press, Washington, DC. https://doi.org/10.17226/25161.

Neff, J.K., Stubblefield, W.A., 1995. Chemical and Toxicological Evaluation of Water Quality Following the Exxon Valdez Oil Spill.

Newman, M.C., McCloskey, J.T., 1996. Time-to-event analyses of ecotoxicity data. Ecotoxicology 5, 187–196.

OECD, 1992. Test No. 306: biodegradability in Seawater. 10.1787/9789264070486-en.

OECD, 2019a. Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures. 10.1787/0ed2f88e-en.

OECD, 2019b. Test No. 203: fish, Acute Toxicity Test. 10.1787/9789264069961-en.

OECD, 2006. *Current Approaches in the Statistical Analysis of Ecotoxicity Data: A guidance to application (annexes to this publication exist as a separate document)*, OECD Series on Testing and Assessment, No. 54, OECD Publishing, Paris. https://doi.org/10.1787/9789264085275-en.

Olsen, G.H., Smit, M.G.D., Carroll, J.L., Jæger, I., Smith, T., Camus, L., 2011. Arctic versus temperate comparison of risk assessment metrics for 2-methyl-naphthalene. Mar. Environ. Res. https://doi.org/10.1016/j.marenvres.2011.08.003.

Parkerton, T.F., Boufadel, M., Nordtug, T., Mitchelmore, C., Colvin, K., Wetzel, D., Barron, M.G., Bragin, G.E., de Jourdan, B., Loughery, J., 2023a. Recommendations for Advancing Media Preparation Methods Used to Assess Aquatic Hazards of Oils and Spill Response Agents. Aquat. Toxicol. 259. https://doi.org/10.1016/j.aquatox.2023.106518.

Parkerton, T.F., French-McCay, D.P., de Jourdan, B., Lee, K., Coelho, G.M., 2023b. Adopting A Toxic Unit Model Paradigm in Design, Analysis and Interpretation of Oil Toxicity Testing. Aquat. Toxicol. 255. https://doi.org/10.1016/j.aquatox.2022.106392.

Peterson, D.R., 1994. Calculating the aquatic toxicity of hydrocarbon mixtures. Chemosphere 29, 2493–2506.

Raimondo, S., Vivian, D.N., Delos, C., Barron, M.G., 2008. Protectiveness of species sensitivity distribution hazard concentrations for acute toxicity used in endangered species risk assessment. Environ. Toxicol. Chem. 27, 2599–2607.

Redman, A.D., Parkerton, T.F., 2015. Guidance for improving comparability and relevance of oil toxicity tests. Mar. Pollut. Bull. 98, 156–170. https://doi.org/10.1016/j.marpolbul.2015.06.053.

Redman, A.D., Mcgrath, J.A., Stubblefield, W.A., Maki, A.W., Di Toro, D.M., 2012. Quantifying the concentration of crude oil microdroplets in oil-water preparations. Environ. Toxicol. Chem. 31, 1814–1822. https://doi.org/10.1002/etc.1882.

Reinert, K.H., Giddings, J.M., Judd, L., 2002. Effects analysis of time-varying or repeated exposures in aquatic ecological risk assessment of agrochemicals. Environ. Toxicol. Chem. 21, 1977–1992. https://doi.org/10.1002/etc.5620210928.

Rico, A., Van den Brink, P.J., 2015. Evaluating aquatic invertebrate vulnerability to insecticides based on intrinsic sensitivity, biological traits, and toxic mode of action. Environ. Toxicol. Chem. 34, 1907–1917.

Sánchez-Bayo, F., 2008. From simple toxicological models to prediction of toxic effects in time. Ecotoxicol 18, 343–354. https://doi.org/10.1007/s10646-008-0290-1.

Salvito, D., Fernandez, M., Jenner, K., Lyon, D.Y., deKnecht, J., Mayer, P., MacLeod, M., Eisenreich, K., Leonards, P., Cesnaitis, R., Leon-Paumen, M., Embry, M., Deglin, S.E., 2020. Improving the environmental risk assessment of substances of unknown or variable composition, complex reaction products, or biological materials. Environ. Toxicol. Chem. 39, 2098–2108.

Singer, M.M., Aurand, D., Bragin, G.E., Clark, J.R., Coelho, G.M., Sowby, M.L., Tjeerdema, R.S., 2000. Standardization of the preparation and quantitation of water-accommodated fractions of petroleum for toxicity testing. Mar. Pollut. Bull. 40, 1007–1016. https://doi.org/10.1016/S0025-326X(00)00045-X.

Sprague, J.B., 1969. Measurement of pollutant toxicity to fish I. Bioassay methods for acute toxicity. Water Res. 3, 793–821. https://doi.org/10.1016/0043-1354(69)90050-5.

Stephan, C.E., Mount, D.I., Hansen, D.J., et al., 1985. Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their uses. PB85-227049. U.S. Environmental Protection Agency, Washington, DC, USA.

Walton, H.E., Davison, J.J., Uzyczak, J., Martin, C., Milliken, P., Kirby, M.F., 2021. The Development of New Toxicity Testing and Approval Processes for Oil Spill Treatment Products in the UK. Int. Oil Spill Conf. Proc. 2021, 687653 https://doi.org/10.7901/2169-3358-2021.1.687653.

Weber, C.I., 1991. Methods for measuring the acute toxicity of effluents and receiving waters to freshwater and marine organisms (fourth edition). United States.

Yu, J., Kindy, M.S., Ellis, B.C., Baatz, J.E., Peden-Adams, M., Ellingham, T.J., Wolff, D.J., Fair, P.A., Gattoni-Celli, S., 2005. Establishment of epidermal cell lines derived from the skin of the Atlantic bottlenose dolphin (Tursiops truncatus). Anat. Rec. A Discov. Mol. Cell Evol. Biol. 287 (2), 1246–1255. https://doi.org/10.1002/ar.a.20266. DecPMID: 16281302.

Zhao, Y., Newman, M.C., 2004. Shortcomings of the laboratory-derived median lethal concentration for predicting mortality in field populations: exposure duration and latent mortality. Environ. Toxicol. Chem. 23, 2147–2153. https://doi.org/10.1897/03-557.