Fisheries and Oceans Canada
Ecosystems and Oceans Science

Pêches et Océans Canada
Sciences des écosystèmes et des océans

Canadian Science Advisory Secretariat (CSAS)

Research Document 2017/064

National Capital Region

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

Julie Adams, Kelli Charbonneau, Danielle Tuori, R. Stephen Brown, and Peter V. Hodson

School of Environmental Studies
Queen’s University
116 Barrie Street
Kingston, ON K7L3N6

December 2017

Foreword

This series documents the scientific basis for the evaluation of aquatic resources and ecosystems in Canada. As such, it addresses the issues of the day in the time frames required and the documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Research documents are produced in the official language in which they are provided to the Secretariat.

Published by:

Fisheries and Oceans Canada
Canadian Science Advisory Secretariat
200 Kent Street
Ottawa ON K1A 0E6

http://www.dfo-mpo.gc.ca/csas-sccs/
csas-sccs@dfo-mpo.gc.ca

ISSN 1919-5044

Correct citation for this publication:

Adams, J., Charbonneau, K., Tuori, D., Brown, R.S. and 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. DFO Can. Sci. Advis. Sec. Res. Doc. 2017/064. xi + 110 p.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ALC	Arabian Light crude oil
ANSC	Alaska North Slope crude oil
aPAH	alkyl polycyclic aromatic hydrocarbons
APHA	American Public Health Association
ASTM	American Society for Testing and Materials
AWB	Access Western Blend
BEWAF	Biologically-enhanced WAF
BTEX	Benzene, toluene, ethylbenzene, and xylenes
BWWAF	Breaking-wave WAF
C	Constant
CB	Commercial blender
CD	Chemical dispersions
CD/E	Chemical dispersion/emulsion
CDO	Chemically-dispersed oil
CEWAF	Chemically-Enhanced Water Accommodated Fraction
CFT	Continuous flow-through
CLB	Cold Lake Blend
CO	Crude oil
CROSERF	Chemical Response to Oil Spills: Ecological Effects Research Forum
DCO	Dispersed crude oil
DCWAF	Dispersed crude WAF
Dilbit	Diluted bitumen
DO	Dispersed oil

DOR	Dispersant-to-oil ratio
DOSS	Dioctyl sodium sulfosuccinate
DSL	Diesel
DWAF	Dispersed WAF
DWH	Deepwater Horizon
EC50	Median effective concentrations
ECCC	Environment and Climate Change Canada
EDCF	Effects Driven Chemical Fractionation
EIA	Environmental Impact Assessment
ERA	Environmental Risk Assessment
GC-FID	Gas Chromatography-Flame Ionization Detection
GC-MS	Gas Chromatography-Mass Spectroscopy
GLP	Good Laboratory Practices
H	Hand mixing
HECEWAF	High Energy CEWAF
HED	High Energy Dispersion
HEWAF	High Energy WAF
HFO	Heavy fuel oil
IFO	Intermediate fuel oils
ISO	International Organization for Standardization
LC50	Median lethal concentrations
LED	Low Energy Dispersion
LEWAF	Low Energy WAF
LOEC	Lowest observable effect concentration
MAC	Macondo crude oil
MDO	Mechanically-dispersed oil
MESA	Medium South American crude oil
MEWAF	Medium Energy WAF
MS	Magnetic stirring
MSB	Mixed sweet blend
MSC	Medium sour composite
MSRC	Marine Spill Response Corporation
N2F0	Number 2 fuel oil
N/A	Not Applicable
NAS	National Academy of Sciences

NOEC	No observable effects concentration
NS	Not Specified
NSC	North Sea crude oil
OAS	Oil accommodated sea water
OECD	Organization for Economic Co-operation and Development
OGE	Oil-gravel effluent
OMAs	Oil mineral aggregates
OS	Orbital shaker
OWD	Oil-water dispersion
OWR	Oil-to-water ratio
P	Pulse or Propeller
PAH	Polycyclic Aromatic Hydrocarbons
PBCO	Prudhoe Bay crude oil
PU	Pump
RSC	Royal Society of Canada
S	Sonicator
SARA	Saturates, aromatics, resins, and asphaltenes
SE	Spiked exposure
SLS	Southern Louisiana sweet crude oil
SNR	Static non-renewal
SR	Static renewal
TEOC	Total Extractable Organic Carbon
TPAH	Total Polycyclic Aromatic Hydrocarbons
TPH	Total petroleum hydrocarbons
TPH-F	Total petroleum hydrocarbons by fluorescence
TUs	Toxic Units
UCM	Unresolved Complex Mixture
UD	Upwelling dilution
US EPA	United States Environmental Protection Agency
UV-VIS	Ultraviolet-visible spectrophotometer
VOC	Volatile Organic Carbons
WAF	Water Accommodated Fraction
WCS	Western Canadian Select
WR	Water recirculation
WSF	Water Soluble Fraction

ABSTRACT

The shipment of diluted bitumen (dilbit) to Canadian refineries and ports is increasing with the expansion of its production in northern Alberta. Dilbit has unique physical and chemical properties relative to conventional crude oils, including rapid weathering and greater adhesion and viscosity that affect its fate, behaviour and effects in marine and freshwater ecosystems. These properties may also interact with commonly used toxicity test protocols to affect measurements of toxicity and change the outcome of ecological risk assessments of dilbit spills. The objectives of this report were: to survey existing methods for oil toxicity tests to identify strengths and weaknesses that could inform the development of standardized methods for testing dilbit; to identify and analyze the advantages and disadvantages of various test methods, recognizing diversity in their objectives; to develop a proposed framework for oil toxicity test methods that links oil type, test species and relevant aquatic environment to the goals of toxicity testing, as a basis for selecting specific test protocols; and to identify relevant research needs.

The Chemical Response to Oil Spills: Ecological Effects Research Forum (CROSERF) recommended standard protocols to prepare Water Accommodated Fractions (WAFs) and Chemically-Enhanced Water Accommodated Fractions (CEWAFs) of crude oil for toxicity tests. The intent was to improve the comparability of toxicity data among studies and their application to Ecological Risk Assessments (ERAs) of oil spills. This survey of peer-reviewed reports of oil toxicity tests with fish catalogued the array of methods actually used. Most reports were published since 2000, and each was assessed to determine if the methods followed CROSERF protocols, provided sufficient information to understand the test results and to support ERAs, and overall, whether there are problems inherent in current practice. CROSERF protocols were also analyzed to determine their suitability for toxicity tests with dilbit.

The survey identified a diversity of toxicity test methods, a lack of detailed reporting of methods, and an inconsistent use of terminology to describe test solutions, all of which hindered comparisons of toxicity among oils, fish species, and environmental conditions. Although most studies were technically sound, they often modified standard methods to suit specific study objectives without providing sufficient details and rationales for the changes reported. Their focus on site-specific conditions limited their applicability to Environmental Impact Assessments (EIAs) and ERAs of oil spills under different scenarios. Fortunately, there is an emerging consensus that test results can be compared when there are sufficient chemical analyses of test solutions to relate measured concentrations of oil in water to measured effects, even when test methods are not described in detail. Chemical analyses have also demonstrated that oil concentrations decline with time during most tests, and that chemical analyses of test solutions do not discriminate between dissolved and particulate hydrocarbons. Therefore, the true toxicity of oil dissolved in water may often be underestimated, adding uncertainty to estimated risks of oil spills.

This review updates the CROSERF protocols with recommendations to overcome specific issues with dilbit. The advantages and disadvantages of various methods are reviewed and a proposed framework is provided on experimental designs, physical and chemical characterization of test oils, preparation of physically- and chemically-dispersed oil, characterizing test solutions, and toxicity testing. Research needs are presented to improve experimental designs, the preparation and toxicity testing of oil solutions, the chemical characterization of test oil and oil solutions, and test requirements that recognize the unique characteristics of dilbit and other crude and refined oils.

Examen des méthodes de mesure de la toxicité pour les organismes aquatiques de la fraction adaptée à l'eau (WAF) et de la fraction adaptée à l'eau améliorée chimiquement (CEWAF) du pétrole

RÉSUMÉ

L’expédition du bitume dilué vers les raffineries et les ports du Canada est en augmentation en raison de l’expansion de sa production au nord de l’Alberta. Le bitume dilué possède des propriétés physiques et chimiques uniques que n’ont pas les pétroles bruts classiques, notamment du fait de son altération rapide et de son adhésion et de sa viscosité plus importantes, qui ont des répercussions sur son devenir, son comportement et ses effets sur les écosystèmes marins et d’eau douce. Ces propriétés peuvent également interagir avec les protocoles d’essai de toxicité couramment utilisés et avoir une incidence sur les mesures de la toxicité, et ainsi modifier le résultat des évaluations du risque écologique des déversements de bitume dilué. Le présent rapport avait les objectifs suivants : relever les méthodes existantes pour les essais de toxicité du pétrole afin de déterminer les forces et les faiblesses pouvant guider l’élaboration de méthodes normalisées en vue de la mise à l’essai du bitume dilué; déterminer et analyser les avantages et les inconvénients de diverses méthodes d’essai, en prenant en compte la diversité de leurs objectifs; élaborer une proposition de cadre pour les méthodes d’essai de toxicité du pétrole qui fait le lien entre le type de pétrole, les espèces d’essai et l’environnement aquatique pertinent et les objectifs des essais de toxicité, qui servira de base pour la sélection de protocoles d’essai précis; et déterminer les besoins pertinents en matière de recherche.

La méthode « Intervention chimique en cas de déversement de pétrole – Recherche sur les effets écologiques (CROSERF) » a recommandé des protocoles standard pour préparer les fractions adaptées à l’eau (WAF) et les fractions adaptées à l’eau améliorées chimiquement (CEWAF) du pétrole brut en vue d’essais de toxicité. L’objectif était d’améliorer la comparabilité des données de toxicité parmi les études et leur application aux évaluations du risque écologique des déversements de pétrole. Cet examen des rapports examinés par les pairs sur les essais de toxicité avec des poissons a catalogué la gamme de méthodes réellement utilisées. La plupart des rapports a été publiée depuis 2000 et chacun d’entre eux a été évalué afin de déterminer si les méthodes respectaient les protocoles de CROSERF, fournissaient suffisamment de renseignements pour comprendre les résultats d’essai et pour appuyer les évaluations du risque écologique et, de façon générale, s’il existe des problèmes inhérents à la pratique courante. Les protocoles du CROSERF ont également été analysés afin de déterminer dans quelle mesure ils sont adaptés aux essais de toxicité du bitume dilué.

L’étude a révélé la diversité des méthodes d’essai de toxicité, un manque de production de rapports détaillés sur les méthodes, et une utilisation incohérente de la terminologie pour décrire les solutions des essais; tout cela a empêché les comparaisons de la toxicité des divers pétroles, des espèces de poissons et des conditions environnementales. Bien que la plupart des études étaient fondées sur le plan technique, elles modifiaient souvent les méthodes standard afin de s’adapter à des objectifs d’étude précis sans fournir suffisamment de détails et de justifications pour les modifications signalées. Le fait qu’elles se concentraient sur des conditions propres aux sites limitait leur applicabilité pour les études d’impact sur l’environnement et les études du risque écologique des déversements de pétrole dans différents scénarios. Heureusement, un consensus apparaît sur le fait que les résultats des essais peuvent être comparés lorsqu’il existe suffisamment d’analyses chimiques des solutions d’essai pour faire le lien entre les concentrations de pétrole dans l’eau mesurées et les effets mesurés, même lorsque les méthodes d’essai ne sont pas décrites en détail. Les analyses chimiques ont également démontré que les concentrations de pétrole diminuent au fil du temps dans la plupart

des essais, et que les analyses chimiques des solutions d’essai n’exercent aucune discrimination entre les hydrocarbures dissous et les particules d’hydrocarbures. Par conséquent, la toxicité réelle du pétrole dissous dans l’eau peut souvent être sous-estimée, ce qui renforce l’incertitude des risques estimés que présentent les déversements de pétrole.

Cet examen met à jour les protocoles CROSERF avec des recommandations visant à surpasser des problèmes précis avec le bitume dilué. Les avantages et inconvénients des diverses méthodes sont examinés et un cadre proposé est fourni sur les plans expérimentaux, la caractérisation physique et chimique des pétroles mis à l’essai, la préparation du pétrole dispersé physiquement et chimiquement, la caractérisation des solutions d’essai et l’essai de toxicité. Les besoins en matière de recherche sont présentés afin d’améliorer les plans expérimentaux, la préparation et la mise à l’essai de la toxicité des solutions de pétrole, la caractérisation chimique du pétrole d’essai et des solutions de pétrole, et les exigences en matière d’essai qui prennent en compte les caractéristiques uniques du bitume dilué et d’autres pétroles bruts et raffinés.

1 INTRODUCTION

Spills of diluted bitumen (dilbit) to aquatic ecosystems followed by its rapid weathering and sinking have raised concerns about whether current toxicity test methods provide reasonable estimates of its hazard and risks. The National Contaminants Advisory Group (NCAG) of Canada's Department of Fisheries and Oceans commissioned this review to determine whether current test methods are adequate for assessing the toxicity of dilbit and other heavy oils, and to propose a reference framework to guide oil toxicity test methods.

Bitumen is extracted from Canada's oil sands and transported from Alberta to markets across North America and to coastal ports. Due to its high viscosity, it must be upgraded to synthetic crude oil or diluted with lighter oils to form dilbit. Upgrading and diluting reduce the viscosity of extracted bitumen to allow flow through pipelines and for loading and unloading rail cars. The physical and chemical properties of dilbit vary considerably depending on the source and extraction method, the type and percent composition of diluent (related to environmental temperatures and mode of transport), and the state of weathering. Diluents are usually natural gas condensates, naphtha or a mix of other light hydrocarbons. Diluents that include synthetic oil are termed 'synbit'.

The growing shipments of dilbit and conventional crude oils have generated concern about the potential impacts on fish and fisheries of spills to marine and freshwater environments. One of the knowledge gaps highlighted is the toxicity to aquatic organisms of dilbit and its components (bitumen and oil-gas condensates). The toxicity of dilbit is likely quite variable, in part due to differences in sensitivity of the wide-array of marine and freshwater species that might be exposed, and in part to the variations in chemical composition of the oil-gas condensates used as diluents, and in the bitumen generated by different extraction processes.

Toxicity will also vary according to the nature of the exposure, i.e., whether dilbit simply floats on the surface of water, is dispersed into small droplets, or coats underwater surfaces. Measuring toxicity requires the preparation of gradients of dilbit concentrations in water, using published methods for creating mechanical and chemical dispersions of oil in water, including mechanical mixing (e.g., Nordtug et al. 2011), addition of chemical dispersants (e.g., Singer et al. 2000), and desorption of hydrocarbons from solid substrates (e.g., Marty et al. 1997). However, the rapid weathering of dilbit during solution preparation can change the concentrations and composition of test solutions, interfering with standard test protocols and preventing unbiased comparisons of toxicity among oils, among species, or among test conditions. The behaviour of dilbit when exposed to air, and the different methods for preparing test solutions, also affect environmental relevance, i.e., the ability to predict the impacts of a dilbit spill under site-specific environmental conditions that might affect rates of weathering and mechanical dispersion in water (e.g., temperature, wind, turbulence).

The need for test methods unique to dilbit illustrates a fundamental conflict between standardized test methods to measure toxicity hazards (i.e., the inherent capacity of petroleum hydrocarbons to cause harm) and the assessment of ecological risk of potential spills and the ecological impacts of actual spills. Because of the wide variations among oils in chemical properties, measurements of toxicity may be biased by differences in their interactions with water during preparation and testing of oil solutions. For example, heavy oils with high viscosities may be more difficult to mechanically disperse than light oils with low viscosities, so that heavy oils may appear less toxic than light oils, despite having higher concentrations of known toxic agents. To establish the least biased estimates of toxicity, high energy mixing or chemical dispersion may be needed to create test solutions with equivalent droplet size distributions and estimates of toxicity that truly reflect the chemical composition of the test oil.

These dispersion methods may not correspond to actual conditions at a spill site, but a single test cannot model the diversity of potential spill scenarios. Scenarios can vary from open-sea surface discharges of finite amounts of oil under storm conditions (rapid dissipation of spilled oil and short-term exposures) to the long-term discharge of large volumes of oil from deep water wellheads (prolonged exposures to surface and particulate oil). Freshwater oil spills are more confined than marine spills and do not dissipate as rapidly, and high energy mixing by turbulent flow in rivers can significantly change the exposure of aquatic species to oil relative to lakes or open oceans. Environmental conditions may also vary widely due to extremes of temperature, salinity, UV-radiation, humic acid content of water, suspended solids, droughts, or flood conditions.

Standardized methods are needed to establish the inherent toxicity of oil, and site-specific tests are needed to assess the effects of the unique environmental conditions at each spill on the exposure of aquatic organisms to oil and on oil toxicity. It is the challenge of Ecological Risk Assessments (ERAs) and Environmental Impact Assessments (EIAs) to integrate toxicity test data from standardized tests with estimated oil exposures of aquatic species under the site-specific conditions of a spill. This review is focused on measuring toxicity without reference to site-specific conditions with the clear exception that variations in standard test methods are needed to reflect: the requirements for toxicity data applicable to marine and freshwater species; the availability of commonly used test species (e.g., cold vs warm-water species); the effects of acute and chronic exposures; and differences among oils in properties that could affect toxicity (e.g., degree of weathering of oil).

The aim of standardizing toxicity test methods is to minimize the variance in results caused by the method itself so that variations in toxicity among tests can be attributed to the different treatments applied (e.g., sources of test oils). Significant issues include the characterization of test solutions, the duration and pattern of exposure, the effect of oil droplets on data interpretation in tests with dispersed oil, and the assumption that oil concentrations in water are correlated to those in tissues (Bejarano et al. 2013; Landrum et al. 2013). Standardization of methods will facilitate comparisons, but laboratory studies are limited in their ability to mimic environmental scenarios (Coelho et al. 2013). Nevertheless, data from standardized toxicity tests can provide insights on the relative risks associated with measured environmental concentrations of oil.

A method that is commonly used to test crude oil, including heavy oil formulations, is the CROSERF (Chemical Response to Oil Spills: Ecological Effects Research Forum) method, which was developed in the United States with Canadian participation. It describes standardized aquatic toxicity test protocols developed to reduce inter-laboratory variability, while at the same time mimicking a ‘realistic’ exposure in which concentrations of waterborne hydrocarbons decline following a single dosing of test solutions. The CROSERF guidance document outlines methods for preparing a Water Accommodated Fraction (WAF) and a Chemically-Enhanced Water Accommodated Fraction (CEWAF) of oil. Canadian researchers applied this method to measure the toxicity of dilbit but found it necessary to modify the recommended protocols to compensate for the rapid weathering of fresh dilbit during testing and for its greater adhesion and viscosity than conventional crude oils (NAS 2016).

To make chemical and toxicological experiments with dilbit more readily comparable among tests and laboratories, more consistent upon repetition, and more useful for environmental risk and impact assessments, there is a need to develop standard reference methods for preparing, analyzing, and testing dilbit WAF and CEWAF. The method should be applicable to most, if not all, petroleum products shipped in Canada. In addition to measuring the aquatic toxicity of ‘fresh’ dilbit products, the method should also include tests of ‘weathered’ products to better understand the toxicological implications of weathering, hours, days, and weeks post-spill.

Nevertheless, oil toxicity tests are highly complex due to the variety of test variables that could affect measured toxicity, but which have not been assessed systematically. Hence, it is difficult to proscribe a single 'standard' method for either conventional oils or dilbit. Therefore, this review presents a proposed framework for testing that allows some flexibility to choose methods that address specific objectives and includes recommendations for chemical analysis of test solutions to facilitate and improve comparisons of results among diverse studies. Throughout the review, research needs are identified in boxes (priority needs in bold), and other recommendations are presented in italics.

1.1 THE PURPOSE AND SCOPE OF THIS LITERATURE REVIEW

This purpose of this review is to promote greater consistency in the preparation, analysis, and aquatic toxicity testing of WAF and CEWAF of dilbit and related petroleum products, more reproducible results, and more meaningful comparisons among test species and petroleum formulations.

The specific objectives of the review are to:

1. Survey the existing methods that have been reported for the preparation, analysis and toxicity testing of WAF and CEWAF of crude oil, refined oils, and dilbit, recognizing the various objectives of the tests, and thus the criteria or requirements of test methods. Given the few reports of dilbit toxicity, this review considers the strengths and weaknesses of test methods for conventional oil products where they inform the development of standardized methods for dilbit;
2. Identify and analyze the advantages and disadvantages of the various methods for preparing, analyzing, and testing solutions of oil in water, recognizing the test objectives;
3. Develop a reference framework for oil toxicity test methods. The proposed framework will link oil type, test species and relevant aquatic environment to the goals of toxicity testing, as a basis for selecting specific test parameters. Method considerations include advice on modifications to be considered for Arctic, freshwater, marine, and estuarine test scenarios, including the Canadian subarctic and temperate zones. Methods will target a suite of relevant, representative Canadian aquatic species (including sensitive life stages) for the different test scenarios. Protocols will include testing fresh and weathered oil and will define and rationalize all preparation and exposure parameters; and
4. Identify research needed to address significant knowledge gaps about methods for determining the fate, behaviour and effects of dilbit in water.

Because of the long history of research on oil spills, the very large number of publications on oil toxicity, and the evolution of technologies for analyzing oil in water, this review was limited primarily to studies with fish since 1990. The period since 1990 encompasses the abundance of research following the 1989 *Exxon Valdez* oil spill and the increasing application of analytical chemistry to characterize test solutions and to relate oil toxicity to specific hydrocarbons. Studies with fish were the most numerous and focused primarily on oil in water, avoiding the added complications of oil in sediments. Because of time limitations, this review did not consider studies with algae and invertebrates. However, most recommendations for test methods with fish, such as the preparation and analysis of test solutions, will also apply to toxicity tests with invertebrates and algae. Recommendations for Good Laboratory Practices (GLP) should be applicable in a general way to any testing protocol. Sediment toxicity test methods for benthic fish, invertebrates, and algae should also be considered in a separate review.

2 BACKGROUND TO DILBIT AND SPILLS OF DILBIT

There have been extensive reviews of the available literature to describe the current state of knowledge on the characteristics, behaviour, and environmental impacts of dilbit compared to conventional crude oils. Two reports stand out in breadth and scope, The Royal Society of Canada (RSC) Expert Panel on The Behaviour and Environmental Impacts of Crude Oil Released into Aqueous Environments (Lee et al. 2015) and the US National Academy of Sciences (NAS) Spills of Diluted Bitumen from Pipelines: A Comparative Study of Environmental Fate, Effects, and Response (NAS 2016).

The RSC Panel reviewed the state of knowledge of oil spills into fresh and marine waters to determine whether it was sufficient to support the development of "…..*effective strategies for spill preparedness, spill response and remediation"* (Lee et al. 2015). The panel identified research needs to support the development of policies and practices in a wide array of disciplines, including: the chemistry and environmental behaviour and distribution of spilled oil; site-specific environmental factors that affect behaviour and distribution of oil; environmental impacts of oil spills; spill response and remediation strategies and their effectiveness; and methods and effectiveness of ERAs and EIAs.

In each part of the RSC report, a specific focus on diluted bitumen (dilbit) explored whether its chemical composition and properties were sufficiently different from conventional crude oils and refined oil products to consider it as a unique product, or simply as a part of a continuum of chemical composition across an array of oil products. In general, the physical and chemical characteristics of various dilbits corresponded to those of medium to heavy oils, but they differed significantly in their behaviour when spilled. Specifically, dilbits weathered at much faster rates than conventional oils, with a corresponding increase in viscosity and density. This behaviour was attributed to the rapid evaporation and dissolution in water of low molecular weight hydrocarbons associated with the diluent component of the dilbit. The results of rapid weathering are an increased resistance of the residual oil to chemical dispersion, an increased propensity to sink when in contact with suspended particulates, a reduction in acute lethality, and a reduction in biodegradation potential due to the high proportions of high molecular weight hydrocarbons. While these properties are similar to those of other heavy oils, dilbit is exceptional in that it resists the formation of water-in-oil emulsions (mousse). Bitumen is a highly weathered oil, and is depleted in alkylnaphthalenes while somewhat enriched in 3- to -5-ring alkyl polycyclic aromatic hydrocarbons (PAH) relative to conventional crude oils. Hence, it has a greater potential for chronic toxicity at equivalent concentrations of total PAH (TPAH). To better understand dilbit behaviour after a spill, and to improve spill model predictions and development of mass balances of spilled oil, the Panel recommended the development of improved methods to measure physical weathering processes such as interaction with particulates and formation of emulsions.

Similar issues and concerns were expressed in the 2016 NAS report (NAS 2016). The committee concluded that the important differences between dilbit and conventional crude oils were "….*in the exceptionally high density, viscosity, and adhesion properties of the bitumen component …….. that dictate environmental behavior as the crude oil is subjected to weathering….".* This dramatic change in properties contributes to the tendency of weathered dilbit to sink or to remain suspended in the water column, primarily because it readily adheres to organic and inorganic particulates. This change in properties with weathering explains much of the fate, behaviour, and distribution of dilbit in aquatic environments following a spill.

The 'binary nature' of spilled dilbit (i.e., the dramatic difference between fresh and weathered states) is highly relevant to the present report, in that changes to physical, chemical and toxicological properties with weathering can confound toxicity tests. Attention must be paid to the condition of dilbit samples when tested and the interaction between test methods (e.g., prolonged mixing that may promote weathering) and measured toxicity. Measured toxicity to aquatic species will depend as much on the exposure to the constituents of dilbit under different weathering scenarios as on their inherent toxicity.

Although ‘petroleum’ encompasses a variety of crude oils and products made from crude oil, the literature reviewed most frequently referred to these products as either ‘oil’ or a specific product (e.g., Alaska North Slope Crude or ANSC), and these general and specific terms will be used throughout.

2.1 PHYSICAL AND CHEMICAL PROPERTIES OF DILBIT PRODUCTS

The physical and chemical characteristics of dilbit exist on a spectrum and are defined by the same terms as conventional crude oils. Dilbit is a blend of bitumen extracted from the oil sands diluted with oil-gas condensates.

According to Yang et al. (2011), the “*oil sands are a mixture of 4% to 6% water, 83 to 85% host sediment (sands, other mineral material), and 10% to 12% bitumen”.* The bitumen component is a heavy oil (API gravity typically <10 ^{o}API; Speight 2006), semi-solid at ambient temperature, and containing a high proportion of asphaltenes and resins and a low proportion of lower molecular weight paraffins and naphthenes compared to conventional oils (Yang et al. 2011). The proportional differences in high and low molecular weight components account for the high viscosity of bitumen relative to other oil products. In response, “upgrading” (synbit) or dilution of bitumen (dilbit) is needed to reduce viscosity and ease handling and transportation.

Dilbit and other oils can be separated into saturates, aromatics, resins, and asphaltenes (SARA). Yoon et al. (2009) reported that a typical Athabasca bitumen sample was 67.97% saturates and aromatics, 15.59% asphaltenes, and 16.44% resins. These proportions vary with the location of the bitumen source and the depth of extraction (Yoon et al. 2009); the composition and degree of biodegradation of bitumen deposits in Western Canada vary among and within deposits (Brooks et al. 1988).

The saturates and aromatics in crude and other oils are normally characterized by gas chromatographic analysis, as presented by Yang et al. (2011). Crude oils typically show a number of sharp, resolved peaks in a chromatogram, resulting from specific saturated alkanes, especially n-alkanes. The chromatogram also normally features a “hump” (or humps) indicating an unresolved complex mixture (UCM) of other compounds, including biodegraded alkanes, higher molecular weight cyclic and branched alkanes, polycyclic aromatic hydrocarbons including alkylated versions, and some hetero-atoms containing alkane derivatives and aromatic compounds. For the crude oil analyzed by Yang et al. (2011), the chromatogram is dominated by the sharp, resolved peaks while the UCM hump can be regarded as a minor feature. In contrast, chromatograms of oil sands bitumen samples consisted mainly of two UCM humps and only minor overlayed sharp peaks, indicating that the individual alkanes, especially lower molecular weight and n-alkanes, were lost from the original crude oils during the formation of the bitumen, likely by biodegradation (Yang et al. 2011). The chromatogram of diluted bitumen resembles a mixture of the bitumen and crude oil chromatograms. The UCM profile is almost identical to that of a bitumen sample, but sharp, resolved peaks are observed on top of the UCM. This is expected where the individual alkanes providing the sharp peaks are from components present in the diluent, which is typically a very light crude oil.

Bitumen has significantly higher levels of aromatic compounds than typical crude oils (Yang et al. 2011). This means that dilbits will also have higher aromatic compositions if the aromatic content of the diluent resembles that of crude oils. Bitumen has minimal levels of low molecular weight naphthalenes (C_0- to C_3-naphthalenes) and for other PAH groups there is a trend among alkyl homologs where C_3>C_2>C_1. This is attributed to the lower susceptibility to biodegradation of the higher molecular weight and more alkylated PAH. Crude oils are usually dominated in total PAH by the naphthalenes, including C_0- to C_3-naphthalenes, and show a trend for other PAH where the highest proportion is C_1 or C_2. The hybrid nature of dilbit is evident from the naphthalene profile which resembles the diluent, while the profile for the other PAH groups typically resembles that of bitumen (Yang et al. 2011). The terpanes and adamantanes of dilbit, known to be resistant to biodegradation, also have a similar profile as bitumen (Yang et al. 2011).

2.1.1 Comparison of properties of dilbit with conventional crude oils

The physical and chemical properties of dilbit vary more widely than conventional crude oils. The characteristics of the blends depend on the source of bitumen and extraction methods, the type of diluent, the ratio of diluent-to-bitumen, and the state of weathering. The API gravity, density, and adhesion of dilbit are distinct from conventional crude oil (Table 1). API gravity (measured in degrees API) is a measure of density relative to water (i.e., the inverse of specific gravity) introduced by the American Petroleum Institute to compare different oil products. The API gravity of fresh water is 10 ºAPI. Light oils have API gravities >31.1 ºAPI, medium from 22.3 ºAPI to 31.1 ºAPI, and heavy <22.3 ºAPI (Speight 2014); the API gravity of heavy fuel oil is 5-15 ºAPI (Table 1). With regard to adhesion, one dilbit product, Access Western Blend (AWB) is similar in adhesion pattern to fuel oils.

Table 1: Physical characteristics of oils (modified from Fingas (2001) and POLARIS (2013)).

Property	**Units**	**Oil Type**							
		Gasoline	**Diesel**	**Light Crude**	**Dilbit[1]**	**Heavy Crude**	**Intermediate Fuel Oil**	**Bunker C**	**Crude Oil Emulsion**
Density	Kg/m^3 at 15˚C	720	840	780 to 880	824 to 941	880 to 1000	940 to 990	960 to 1040	950 to 1000
API Gravity	ºAPI	65	35	30 to 50	18 to 39	10 to 30	10 to 20	5 to 15	10 to 15
Viscosity	mPas at 15˚C	0.5	2	5 to 50	270.5* to 265,263 **	50 to 50,000	1,000 to 15,000	10,000 to 50,000	20,000 to 100,000
Flash Point	15˚C	-35	45	-30 to 30	<-35**^{m} to 58*^{m}	-30 to 60	80 to 100	>100	>80
Solubility in Water	ppm	200	40	10 to 50	-	5 to 30	10 to 30	1 to 5	-
Pour Point	˚C	NR	-35 to -1	-40 to 30	-30**^{m} to 15**^{m}	-40 to 30	-10 to 10	5 to 20	>50
Interfacial Tension	mN/m at 15˚C	27	27	10 to 30	27*^{m} to 150*^{m}	15 to 30	25 to 30	25 to 35	-

[1] Values provided include weathered dilbit from tests; NR= not relevant; *Calculated for AWB; **Calculated value for CLB; *^{m}Measured value of AWB; **^{m}Measured value of CLB

Dilbits also differ from conventional crude oils in their SARA composition. Yang et al. (2011) surveyed Alberta's oil sands, dilbit, synthetic oil, and conventional crude oil and determined that bitumen and dilbit can be distinguished from most conventional crude oils based on their composition, including the distribution of PAH homologs, the n-alkanes, the UCM, and terpane and adamantine biomarkers (Yang et al. 2011). In particular, the proportional increase in PAH homologs with degree of alkylation differs from the typical bell-shaped distribution of conventional crude oils, with a higher proportion of more alkylated congeners in dilbit (Yang et al. 2011). The most abundant PAH in conventional crude oils are the C_0 to C_4 alkyl naphthalenes. They are also abundant in dilbit products due to the addition of a diluent, but a relatively small proportion of TPAH in bitumen. POLARIS (2013) reported that the PAH fraction in dilbit is similar to crude oils, and less than intermediate fuel oils (IFOs).

As with conventional crude oils, dilbit undergoes physical and chemical changes with weathering processes following a spill. Low molecular weight volatile compounds are lost rapidly and adhesion changes the most in the first 24 hours after a spill, with little change in the following 280 hours. In contrast, there is a linear correspondence between adhesion and weathering time for light crude oils (Hollebone 2015 and ECCC unpublished data referenced in Lee et al. 2015). The mechanisms of weathering, the rate of change, and the degree of change are determined by the composition of the oil and by the conditions of the environment. However, the weathering of dilbit is distinct from other oils. Weathering of dilbit occurs on a shorter time scale due to the rapid loss of the diluent. This also means that changes in physical characteristics of dilbit with weathering, such as increased density, occur more rapidly than for crude oils. The proportion of oil lost in weathering is less than for other oils, and the final weathered state is similar to the original bitumen component (NAS, 2016). Due to the change in the characteristics of dilbit when weathered, the NAS panel (2016) recommended that the United States Coast Guard change their classification of dilbit products to highlight its tendency to sink.

2.2 REVIEW OF SPILLS OF DILBIT PRODUCTS

Dilbit products are transported across Canada via pipelines, truck, and rail, and from the coast via tanker. With proposals for new pipelines to transport dilbit from Alberta to international markets via the west and east coasts, there are increasing risks of spills to freshwater ecosystems. The mode and route of transportation contribute significantly to the potential for dilbit spills, including the duration of release, total volume released, and feasibility of rapid response and clean-up. Dilbit products flow more readily at lower temperatures and spread further than undiluted bitumen or heated bitumen that would rapidly cool when spilled. Evaporative weathering, dissolution of water soluble components, uptake of solid matter and water, biodegradation, emulsification, and photo-oxidation will also influence the fate of spilled dilbit (Dew et al. 2015).

There is debate over the sinking behaviour of dilbit products. As with Group III and IV petroleum-based and synthetic oils, sinking of dilbit depends on environmental conditions (POLARIS 2013), and on chemical composition, density, volatility, viscosity, and pour point. The wax and asphaltene content may influence the ability to form water-in-oil emulsions, and the asphaltene and resin content influences the formation of tar balls (POLARIS 2013). Given these characteristics, dilbit may sink or form tar balls during solution preparation for toxicity tests, depending on the mixing protocol. There are no reports yet that these behaviours have affected laboratory toxicity tests, but dilbit has been observed to sink following a spill into Talmadge Creek and the Kalamazoo River, Michigan. On July 25, 2010, a two meter rupture in the Enbridge Line 6B pipeline released 27,000 barrels of dilbit into Talmadge Creek, Michigan (Brooks 2014). The release occurred at the end of a batch of Western Canadian Select (WCS) and the beginning of a batch of Cold Lake Blend (CLB); the released oil was estimated to be 77.5% CLB and 22.5% WCS (Enbridge Energy 2013).

A summary of the release was provided by Brooks (2014):

- Dilbit spilled on land, where saturated soils from heavy rainfall allowed it to travel to Talmadge creek and then to the fast-flowing and turbid Kalamazoo River.
- During the flood, oil slicks were broken up by vegetation and in the open river channel there was entrainment and sedimentation. When the water level decreased below the floodplain to pre-flood levels, there was stranded oil on the vegetation, in oil-filled macropores (defined by

Brooks (2014) as *"root traces and animal burrows that can potentially trap oil and contaminate the subsurface"*), and on land. Oil was deposited in sediments of river channels and sheens formed on the water surface when it floated free.

- The faster-than-expected evaporative weathering during travel and incorporation of particulate matter, both on land and in the water column, resulted in multiple environmental fates of dilbit in water, including floating, submergence, and sinking.
- Receptors in multiple habitats of the river system were disturbed.

After the first few weeks of clean-up of floating oil, the focus shifted to the recovery of submerged oil (FOSC 2016) indicated by the presence of oil globules resurfacing spontaneously or when sediments were disturbed (Enbridge Energy 2013). Estimates of the amount of oil remaining in the sediments differ between Enbridge and the USEPA and range from 19,000 to 181,000 gallons (FOSC 2016).

In Canada, the largest inland spill of a bitumen product in Canada to date was the Nexen Energy pipeline rupture in Long Lake, AB on July 15, 2015. A pipeline fail-safe malfunction released 31,000 barrels of an emulsion of bitumen (approximately 33% oil), water and sand (Prystupa 2015) that covered approximately 21,900 m^2 (Nexen 2015). The excavation of the released bitumen from the site was completed on April 19th, 2016, but there have been no recent news articles or reports of impact assessments.

Similarly, the fate and effect of approximately 1,572 barrels of heavy oil (not dilbit) spilled into the North Saskatchewan River from a Husky Energy pipeline between 8 PM July 20th, and 6 AM, July 21st, 2016 (Nickel & Williams 2016) have not yet been reported publicly. The type of oil spilled was HLU Blended LLB Heavy Crude Oil (Saskatoon Star Phoenix: Crews unlikely to recover all the spilled crude: Ministry of Environment). The fast currents of the North Saskatchewan River moved the oil rapidly downstream, as far as 500 km by July 28th (Cotter 2016), and two municipalities were forced to shut down their drinking water intakes (McSheffrey 2016). Because the oil spill occurred during a high rain period, it was predicted that the oil would be deposited on shoreline grasses and vegetation (CBC 2016a). The amount of debris in the water also increased the potential for oil interactions with suspended debris in the water column (McSheffrey 2016), but as of September 2016, there were no detailed reports of sediment contamination or formal EIAs. Nevertheless, Ricardo Segovia, a hydrogeologist, reported that oil had settled to the bottom of the river (CBC News: More calls for transparency in the aftermath of Husky oil spill), which would mirror the experience at the 2010 Kalamazoo spill.

As of July 30th, there was no floating oil remaining on the water, and the oil on the banks was being washed into the water for collection and clean-up. The residual oil was described as "very thick and viscous". Ed Owens reported that the oil spilled "is a floating oil" and all of the samples taken so far do not indicate oil associated with bed sediments, although no description of the types of samples or sampling locations was provided (CBC 2016b).

2.2.1 Review of exposure scenarios in dilbit spills

As with other types of oils, floating dilbit can be collected from the water's surface, presenting a lower exposure to aquatic species. There is increased concern about the impacts of spills when oil submerges within the water column and persists. Dilbit products are of particular concern because of its rapid weathering and sinking behaviour (Dew et al. 2015), as evident in the 2010 Kalamazoo River spill. Specific conditions for the sinking behaviour of dilbit in fresh water were identified in laboratory tests (review, Dew et al. 2015). In one report, dilbit was added to fresh water at 15˚C in a laboratory tank, with a light wind (S.L. Ross 2012) and weathered for 11 days to a density slightly less than that of fresh water; it showed no signs of sinking. Under the same conditions, but with exposure to UV light, dilbit droplets were observed in the top 10 cm of the water column, likely due to chemical changes that affected surface tension. However, there was no sign of dilbit sinking. In contrast, under similar conditions, the presence of organic or inorganic particulates in water was associated with precipitation of dilbit to the bottom of test vessels (GOC 2013), corresponding to its behaviour in the Kalamazoo River (Dew et al. 2015). Flume tank studies of AWB added to filtered brackish water (< 17 ppt salinity) at 19˚C demonstrated that weathering in low energy conditions void of suspended particulates > 5 µm resulted in sinking oil droplets within 7 days (King et al. 2014). The authors speculated that weathered AWB would sink in fresh water under similar conditions.

Given the observed sinking behaviour of dilbit, spilled dilbit will persist if not recovered immediately, increasing the risk of chronic oil exposures of benthic organisms. In marine environments, conventional oil typically floats on water, spreads over the surface and may be mechanically-dispersed by wind and waves, or recovered. Dilbit should also float because the density of sea water is greater than that of fresh water. Nevertheless, turbidity in river estuaries and in near-coastal waters could provide a source of particulates that would facilitate sinking. Given the potential for sinking, research is needed to develop standardized methods to assess the toxicity of dilbit incorporated into sediments.

The behaviour of spilled dilbit may also affect the efficiency of chemical dispersion following marine spills (Lee et al. 2015). Weathered oils are less susceptible to chemical dispersion than fresh. Because of rapid weathering, the window of opportunity for chemical dispersion of dilbit will be shorter than that of conventional crude oils. The interval during which dispersants can be used will also be determined by site-specific environmental conditions (e.g., temperature, wind speed, wave turbulence) that affect the rate of dilbit weathering.

2.2.2 Linking dilbit behaviour observed in spills to toxicity testing methods

Although there is limited literature on field studies of dilbit spills, it is clear that the environmental behaviour and fate of dilbit change rapidly with weathering and interaction with particulates. These changes are highly relevant to toxicity testing, and comparisons of toxicity between chemically-dispersed and mechanically-dispersed dilbit. Published protocols for preparing test solutions involve mixing oil and water for 18 hours (WAF), and before adding dispersant (CEWAF) (e.g., Wu et al. 2012). The increased viscosity of dilbit over 18 hours of stirring may severely reduce the partitioning of hydrocarbons from dilbit to water and the effectiveness of chemical dispersants. It is critically important to recognize these characteristics of dilbit products when designing oil-water mixing because they will affect the results of toxicity tests and the perception of risk relative to spills of conventional oils.

Research Need

- Research is needed on the optimum mixing times for the most effective chemical dispersion and the most repeatable tests of dilbit toxicity, a recommendation that applies to all oils.
- Develop standard methods to assess the toxicity of dilbit incorporated into sediments.

3 BACKGROUND TO TOXICITY TESTS WITH OIL

The 2015 report of the Royal Society of Canada (Lee et al. 2015) included a review of the effects of oil exposure on aquatic biota, factors that affected measurements of toxicity, and the utility of published toxicity data for ERAs and EIAs. Recognizing that toxicity tests are meant to be controlled experiments in which the only variable is a graded series of oil concentrations, the report identified a series of assumptions that, if violated, would reduce the accuracy of toxicity estimates and their application to ERAs and EIAs (Box 1). These assumptions were rarely tested or discussed in the literature surveyed, indicating that most published toxicity data include significant uncertainties that reduce their accuracy and utility for ERAs and EIAs.

Box 1. Assumptions about oil toxicity tests

(adapted from Lee et al. (2015) p 171)

- All test variables are constant except a graded series of oil concentrations
- Concentrations of hydrocarbons in test solutions increase proportionately with the amount of oil added
- Hydrocarbons are freely dissolved in test solutions
- Concentrations are known and constant throughout the test
- The relative proportions of hydrocarbons remain constant over time and among test concentrations
- Test organisms respond only to petroleum hydrocarbons

Among the key issues and research needs identified were:

- Time-varying concentrations: significant and rapid changes in the concentrations and composition of petroleum hydrocarbons during toxicity tests (see Section 4.4.5);
- A lack of data on uptake-depuration constants for petroleum hydrocarbons to support modelling of internal dose associated with time-varying concentrations of oil;
- A focus on effects of acute (24-96 h) exposures to oil and fewer data on prolonged or chronic exposures typical of spill scenarios such as the Deepwater Horizon (DWH) Oil Spill or pipeline oil spills;
- Few data on prolonged or delayed responses to short-term exposures, impacts on physiological performance (e.g., swimming performance as a key element of feeding, predator avoidance, and spawning migrations), and impacts on long-term growth, reproduction, and rates of disease and survival;
- A need for detailed chemical characterization of test solutions to describe time-varying concentrations of oil and to identify individual components that may be responsible for acute and chronic toxicity;
- A need for rapid and cost-effective analytical methods to support detailed chemical characterizations of:
  - petroleum hydrocarbons in test solutions;
  - components of chemical dispersants (e.g., dioctyl sodium sulfosuccinate (DOSS) in Corexit 9500) to understand the role and relative importance of dispersant toxicity or interactive toxicity of dispersant-oil mixtures;
  - droplets in test solutions to describe the distribution of droplet sizes associated with different methods of preparing solutions of oil in water;

- A need for sampling methods that discriminate between the ‘dissolved’ and ‘particulate’ phases in test solutions to support an understanding of the relative bioavailability and toxicity of dissolved and liquid phase hydrocarbons;
- Experimental designs that represent the wide array of exposure scenarios in marine and fresh waters, in lakes, rivers and wetlands, and in Arctic waters subject to brief or continuous spills; and
- A recognition that the results of toxicity tests are highly dependent on the methods applied for mixing oil and water or oil, water and chemical dispersants. Methods developed to represent one site-specific exposure scenario may not be applicable to the myriad of other spill scenarios. The characteristics of each oil are unique, as are the ecological characteristics and environmental conditions where the oil is spilled.

The issues identified by the RSC panel report are revisited in more detail in Chapter 4 of this report.

3.1 BRIEF INTRODUCTION TO AQUATIC TOXICITY TESTING

Aquatic toxicity tests assess the hazards of chemicals or mixtures of chemicals in water. They provide data for predicting the effects of environmental exposures and comparing relative toxicities among oils, aquatic species, and test conditions. According to the *Method Development and Applications Section* of Environment and Climate Change Canada (Report EPS 1/RM/9 with May 2007 amendments) a “*toxicity test is a determination of the effect of a material on a group of selected organisms under defined conditions*.” The ideal test will cover a range of concentrations, generate a gradient of effects, and describe a repeatable exposure-response relationship.

The important components of test protocols and the need for standardization were first noted by Doudoroff et al. (1951) for toxicity testing of industrial effluents. The textbook *Aquatic Toxicology* (Rand 1995) contains extensive reviews of test methods and of the variables to be considered in their design and execution (Cooney 1995; Rand et al. 1995; Sprague 1995; Ward 1995). Many of the principles and test conditions that influence the outcome of toxicity tests using fish are also reviewed in detail by Sprague (1969; 1970; 1971). The test conditions recommended by these authors constitute ‘best practice’ and are reflected in published test protocols (e.g., Environment Canada 1998; 2007; 2011).

Standardizing test methods allows multiple practitioners with varying skill levels to generate comparable and consistent data for routine monitoring (Davis 1977). The process for standardizing test methods was described by Davis (1977), and involved a balance between the strict outline of the details and the practicality and relevance of the test outcomes. Methods can be too prescriptive, making their application tedious, impractical, inflexible in new circumstances, or insufficiently detailed to enable replication. The advantages and disadvantages of standardization vary with the objectives of the test. For example, detailed prescriptive methods are beneficial for regulatory approval of chemicals but may be too restrictive for toxicity tests used in research or irrelevant to the environmental conditions of an actual spill (Davis 1977). Rand (1995) outlined criteria that may be used in evaluating a study to determine if it should be a standard test, including whether the proposed test is: widely accepted, repeatable, easy to conduct, economical, sensitive, and can be used in an ERA.

The development of test methods and guidelines increased the reproducibility and comparability of test results. Such guidelines are established and disseminated by several groups including Environment and Climate Change Canada (ECCC), the United States Environmental Protection Agency (US EPA), the Organization for Economic Co-operation and Development (OECD), the

International Organization for Standardization (ISO), the American Public Health Association (APHA), and the American Society for Testing and Materials (ASTM).

This review focuses on aquatic toxicity test methods using fish, for which there are numerous methods and guidance documents. For example, ECCC's Biological Test Methods include an Acute Lethality Test Using Rainbow Trout (EPS1/RM/09), an Acute Lethality Test Using Threespine Stickleback (*Gasterosteus aculeatus*) (EPS1/RM10), a Reference Method for Determining Acute Lethality of Effluents to Rainbow Trout (EPS1/RM13), a Test of Larval Growth and Survival Using Fathead Minnows (EPS1/RM22), and Toxicity Tests Using Early Life Stages of Salmonid Fish (Rainbow Trout) (EPS1/RM28).

3.2 STANDARD METHODS EMPLOYED IN OIL TOXICITY TESTING

The design of toxicity tests will depend on the purpose and goal of the test. One such goal is to provide toxicity data of sufficient quality to support ERAs of oil spills, particularly of dilbit, to aquatic ecosystems under different environmental conditions. Standard test methods can also be applied to support remediation, impact assessment, research, and regulatory and legal requirements.

A challenge with the development of standard methods for oil is that many methods were developed for single compounds or for liquid effluents containing water-soluble compounds. However, crude and refined oils are complex mixtures of compounds with variable water solubility and volatility. Further, the composition of each oil is unique, depending on its type, source and additives (CONCAWE 1993). It is imperative that methods developed for water soluble substances should be modified to reflect the properties of oil (Rufli et al. 1998).

Girling (1989) reviewed various methods to generate aqueous solutions for tests of oil products. The review described two types of solutions; 'whole' media solutions that contain bulk oil, which included chemical dispersion/emulsion, high energy dispersion and low energy dispersion, and 'fractional' media solutions in which the aqueous phase contains dispersed and dissolved oil. Methods for preparing solutions varied widely, and included oil-to-water ratios (OWR) in sea water from 1:18 to 1:1000, mixing durations from 4 to 240 hours, and settling times of 0 to 4 hours. Girling (1989) identified several factors that influence the composition of the test media, including the properties of the product, the mixing environment, quality of the aqueous phase, properties of the components in the oil mixture, and the OWR.

Similarly, LaRoche et al. (1970) called for standardizing diluent water temperature, pH, salinity (synthetic sea water prepared with tap water), dissolved oxygen, mixing methods, and the use of reference toxicants to allow comparisons among studies of dispersed oil. Additional factors identified that affect the composition of test solutions were oil type, mixing time, and temperature (Maher 1982), and the OWR and presence of particulate oil (Maher 1986). A large amount of research on dispersants, oil, and dispersed oil toxicity has since been conducted with different toxicity test methods and different analytical standards. However, the properties of test solutions listed above are rarely measured, which limits the comparability of results among labs, among oils, and among species (Singer et al. 2000, 2001a). Thus, measured toxicity appears to be largely dependent on the characteristics of the test. Given the complexity of oil and the number of factors in the design of toxicity tests, variation among toxicity test results is not surprising (Singer et al. 2001a).

3.3 NOMENCLATURE

Only a small portion of the studies reviewed in this report measured the toxicity of multiple oils. Hence, comparisons of toxicity among oils, species, or environmental conditions involve comparisons among publications that employed different methods. A primary impediment to such comparisons is a lack of standardized terminology that corresponds to the methods applied, and there is a critical need to report experimental details using standard terminology consistently. The terminology applied throughout this review is consistent with CROSERF (Section 3.3.3). Among the 144 publications reviewed, there were 54 distinct terms or descriptions of the 223 solutions and/or methods reported, including modifiers of commonly used terms, most of which are variations on a few main terms (Table 2). The frequency with which the main terms were used in publications is illustrated in Figure 1.

Table 2: List of commonly used terms

Terminology	Description	Related terms reported in the literature
Water Accommodated Fraction (**WAF**)	Generated by low energy mixing of oil floating on water, with little disturbance of the surface slick. Test solutions contain dissolved hydrocarbons and potentially a small fraction of particulate oil	LEWAF, MEWAF, HEWAF – WAF prepared by low, medium or high energy mixing BEWAF – biologically-enhanced WAF BWWAF – breaking-wave WAF Burnt Crude oil WAF Crude Oil WAF
Water Soluble Fraction (**WSF**)	WAF that has been filtered or treated to remove all oil droplets and contains only 'truly soluble fractions of each test oil' as described by Anderson et al. (1974).	Diesel WAF WAS – WSF of diesel WCO WSF – whole crude oil WSF WSFd – WSF of dilbit WSF doses
Chemically-Enhanced Water Accommodated Fraction (**CEWAF**)	Generated by mechanical and chemical dispersion of floating oil. Test solutions contain dissolved hydrocarbons and a significant amount of particulate oil.	WAF of oil and dispersant HECEWAF – high energy CEWAF
Dispersed Oil	Particulate oil dispersed mechanically or chemically	DO – dispersed oil WAF DO – WAF of dispersed oil DCO – dispersed crude oil DCWAF – dispersed crude WAF DWAF – dispersed WAF CD – chemical dispersions CDO – chemically-dispersed oil MDO – mechanically-dispersed oil

The potential for confusion was evident throughout this literature review. Many nuances in terminology were used to describe oil/water solutions and/or the methods used to prepare the solutions. The terms used did not consistently refer to the oil/water test solution itself, the method used to prepare the solution, or either. For instance, when a given paper denotes an oil-water mixture as a WAF, one might assume the terminology refers to the product of low energy mixing on a magnetic stir-plate, as defined under CROSERF (Aurand & Coelho 2005). However, WAF sometimes refers to the solution created by dilution water flowing over an oiled substrate (i.e., Adams et al. 2014a).

There were similar issues with 'WSF' which often refers to oil/water solutions ostensibly similar in composition, but prepared by very different methods. For example, Akaishi et al. (2004) used WSF to refer to solutions prepared by adding oil to slowly-stirred water with mixing on a magnetic stir-plate, following Anderson et al. (1974). Alternatively, WSF was generated by dilution water flowing through a polyvinyl chloride column filled with ceramic beads pre-soaked in oil (Kennedy & Farrell 2005). Conversely, some publications used different terms to describe the same preparation method or test solution. For example, Santos et al. (2013) and Mattos et al. (2010) generated WSF and WAF, respectively, although both followed the mixing protocol outlined by Singer et al. (2000). Without chemical characterization, it is difficult to determine if solutions from two different studies really are equivalent. WSF has been differentiated from WAF by filtration to remove particulate oil, but the verification that droplets were absent by chemical analysis or filtration (Anderson et al. 1974; Carlsson et al. 2014) was rare. The presence of particulate oil skews chemical analyses of test solutions and interpretation of toxicity. For this reason, CROSERF adopted the WAF terminology because it does not imply that particulate oil or oil micelles are absent (Faksness et al. 2008). This report considers WSF to be equivalent to WAF, and where applicable there is a description of the presence of particulate oil.

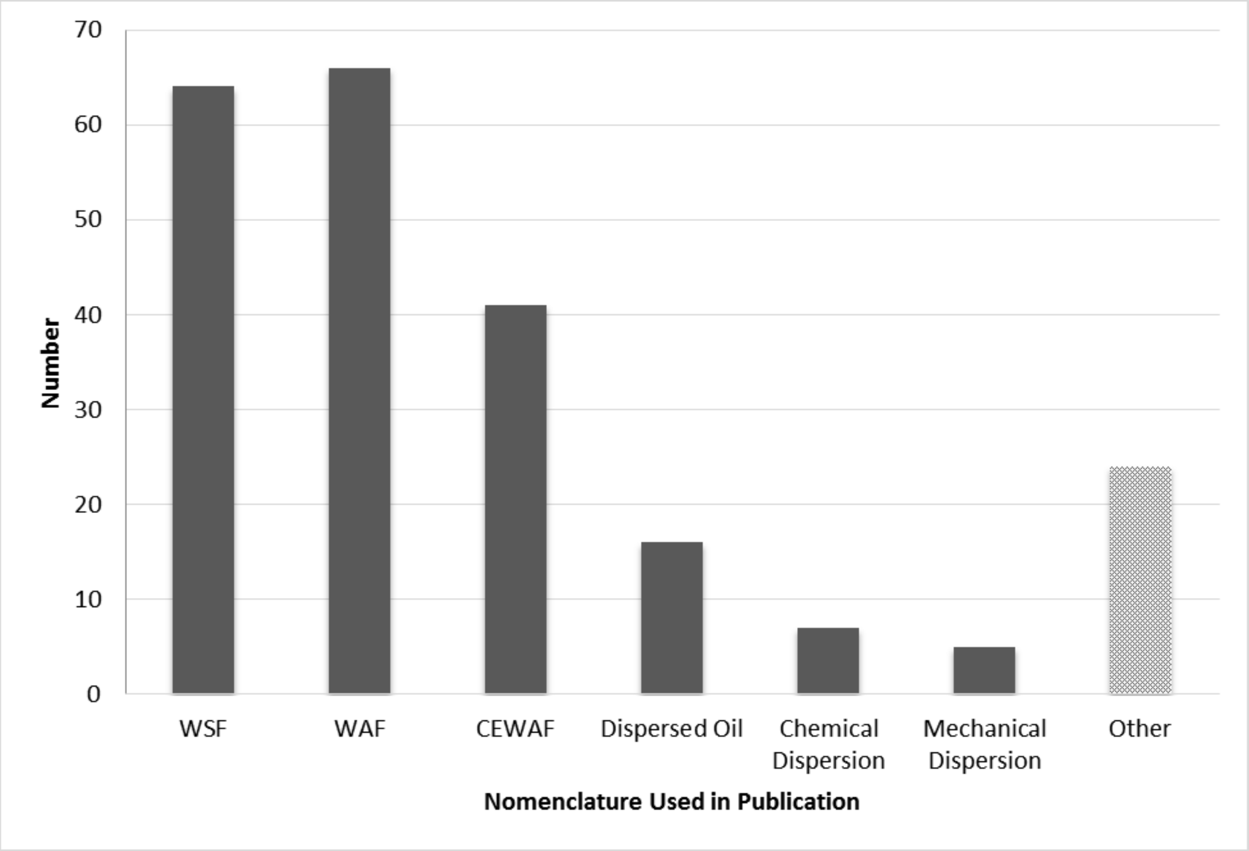

Figure 1: The number of terms describing oil/water solutions, and/or methods to prepare those solutions, for toxicity tests (N=223) (from Table 2). ***Other*** *refers to terms used three times or fewer (listed in Table 3).*

Table 3: List of terms that occurred three or fewer times in the literature review and classified as ***Other*** *(Figure 1). Due to the similarity of the descriptors, crude oil included 'crude oil (CO)'.*

Number of term uses = 1	
• Effluent from columns of gravel coated with stranded oil • Effluent from oil-coated gravel column • Effluent from oiled rock columns • Emulsified oil • Emulsions of oil and dispersant • Floating oil • Oil accommodated sea water (OAS)	• Oil coated gravel and effluent • Oil in dispersant media • Oil only media • Oil-water dispersion (OWD) • Physically- dispersed water accommodated fraction (Physically-dispersed WAF) • Water percolated through oil-coated gravel
Number of term uses = 2	
• Oil contaminated gravel	
Number of term uses = 3	
• Crude oil	• Oil-gravel effluent (OGE) • Treatment water from oiled gravel

Some practitioners also modified commonly used terms such as WAF to convey additional information about oil dispersion, oil type, mixing method, mixing energy, etc. For example, '**WSFd'** referred to the water-soluble fraction of dilbit (Alderman et al. 2016). Alternatively, '**WSFD**' or '**WSD'** described the water-soluble fraction of diesel (Simonato et al. 2008), but Hedayati and Jahanbakhshi (2012) used '**WSD'** to describe 'water-soluble doses' of hydrocarbons. Clearly, inconsistency in the use of terms makes it difficult to infer important details about the method and the nature of the test solutions, as does too little information (e.g., using the term '**dispersed oil'** with no additional descriptors).

About 11% of the methods were classified as **Other** (terms that could not be grouped into the categories in Table 1 and used by three publications or fewer), reflecting the inconsistency in the naming of test solutions and/or preparation methods.

Although many terms are equivalent, this similarity cannot always be inferred from the nomenclature alone. There is a critical need for standardized terminology that clearly conveys accurate information about the test solution and/or the preparation method. Standard terminology should acknowledge the original or published definition of the term, ensure clear communication of the term as a descriptor of either the test solution or the method preparation (or both), and use new acronyms only when not available from other applications. The proposed framework for this review (Chapter 5) uses the WAF and CEWAF terms defined by CROSERF, modified where needed to indicate the energy applied to mixing. It should also be recognized that terms such as 'mixing energy' are relative (e.g., low vs high) or operationally-defined (i.e., stir bar speeds to generate a specified vortex). No technologies were reported for quantifying mixing energy in laboratory containers or at sites of oil spills (see section 3.2.2).

To reduce confusion, each published method reviewed in this report was categorized according to mixing energy, addition of a chemical dispersant, and differentiation between addition of oil to water or a substrate (Table 4). The categories of methods to prepare test solutions in Table 4 not only improved consistency, but also identified which methods were used most often. The majority of the 223 methods surveyed were characterized as **Physical/mechanical dispersion** in Table 4. In contrast, Figure 1 demonstrates that the most commonly used term was WSF, which implies mixing with no vortex; WSF studies should be classified under **Undispersed**. However, the **Physical/mechanical dispersion** group of Table 4 includes 27 methods designated as WSF by the authors when in fact they used sufficient mechanical mixing to produce a dispersion. Similarly, 12 methods described as generating WAF were categorized as **Undispersed** as they followed the CROSERF WAF protocol specifying low-energy mixing to avoid vortex formation (Aurand & Coelho 2005).

The second most common characterization was **Chemical dispersion** of test solutions prepared with chemical dispersant plus mechanical mixing, although it was not always clear that the protocol followed CROSERF guidelines. A smaller number of methods were characterized as **Oiled substrate**, which included a diverse array of terms, including WAF, oil-gravel effluent (OGE), effluent from columns of gravel coated with stranded oil, etc. Oiled substrate methods simulate the impacts of oil on embryos in spawning shoals. These could be described as aquatic toxicity tests because fish are exposed to water that was in contact with oil (WAF), or sediment toxicity tests because toxicity can be related to the concentration of oil in sediments. A small portion of methods were not detailed enough to assign them to a category in Table 4.

Table 4: List of methods used to prepare oil/water solutions for toxicity testing. These characterizations are general descriptions and may or may not agree with the nomenclature used to describe the method of preparation in each publication (Figure 1).

General Description of Method Used	Key Method Details	Number	% of Total Methods
Undispersed	Floating oil, no vortex	30	13.5
Physical/Mechanical dispersion	Vortex or breaking of oil into droplets, no differentiation of dissolved and particulate oil	83	37.2
Chemical dispersion	Addition of a chemical dispersant, includes physical/mechanical dispersion	66	29.6
Oiled substrate	Addition of oil to a substrate followed by desorption	19	8.5
Insufficient information	Publication did not provide enough detail to classify under the general descriptions	25	11.2
Total:		223	N/A

3.4 THE CROSERF PROCESS

In response to the limitations of reported oil toxicity tests (section 3.0), the CROSERF process was initiated by the State of California, and later the Marine Spill Response Corporation (MSRC) to improve the comparability of test results and to better predict the toxicity of dispersants and of chemically-dispersed oil. This major collaborative effort among academia, government, and industry was initiated in 1994 to generate standard oil toxicity methods. The Technical Report by Aurand and Coelho (2005) outlined the goals and findings of the CROSERF initiative and is summarized below.

The specific objectives were to resolve scientific issues related to dispersant use, encourage standardization, foster cooperative efforts, integrate appropriate laboratory data into the dispersant decision process, and encourage the development of risk assessment protocols. The difficulties in interpreting data on dispersant and dispersed oil toxicity stemmed from the methods used for laboratory studies and the relationship between laboratory studies and environmentally-relevant exposures. As noted by Aurand and Coelho (2005), analysis and comparison of results were inhibited by insufficient characterization and reporting of hydrocarbon concentrations in test solutions, unknown effects of temperature on toxicity, and a lack of information about effects at environmentally-realistic concentrations. A major concern was that many studies often tested unrealistically high concentrations of oil and longer exposure durations than might be experienced at a spill, based on a scenario of an offshore marine oil spill.

The Forum produced a set of methods that, when standardized, reduced inter-laboratory variability in oil toxicity tests. The components of toxicity tests that were identified as the highest priority for standardization by Singer et al. (2001a) were the methods to prepare the test solutions, exposure regimes, and the analytical measurement of test solutions.

These priorities are still relevant, although the scenarios considered in toxicity testing have expanded considerably to include spills to Arctic, sub-Arctic, and freshwater ecosystems (lentic and lotic). Spill events, such as the DWH well blowout, provided significant funds for research that fed into hazard and risk assessments (Redman & Parkerton 2015). Unfortunately, the new research did not always follow the recommended methodologies for oil toxicity testing and generated data that were not readily comparable to those from other studies. With 15 years' experience, shortcomings of the CROSERF protocols have also been identified, and the present review and others have identified a need to increase the quality of experimental designs to assess oil as a complex mixture.

3.4.1 Guidance documents for improving oil toxicity studies

There are potential issues with the quality of toxicity test data related to: precision; accuracy; statistical certainty; interpretation of results; comparability of studies among different oils, test conditions, species, and endpoints; appropriateness of test conditions; and extrapolation of laboratory-derived data to site-specific conditions and exposures. Regulators require that toxicity tests produce clearly defined and repeatable exposure-response relationships with adequate characterization of the exposure to assess risk of contaminants with predicted or measured environmental concentrations (Harris et al. 2014) (Box 2). Establishing causation is more attainable with individual chemicals or simple mixtures, and becomes more challenging with increasing complexity of test solutions. Interpreting oil toxicity tests is complicated because most of the components of oil are unknown and of very low water solubility (Landrum et al. 2013). Comparisons are further confounded by multiphase test solutions in which particulate and dissolved phases may contribute to the toxicity of the mixture (e.g., Sørhus et al. 2015).

Box 2. Principles of Sound Ecotoxicology
(Harris et al. 2014)

- Adequate planning and good design of a study are essential
- Define the baseline
- Include appropriate controls
- Use appropriate exposure routes and concentrations
- Define the exposure
- Understand your tools
- Think about statistical analysis of the results when designing an experiment
- Consider the dose-response
- Repeat the experiment
- Consider the confounding factors
- Consider the weight of evidence
- Report findings in an unbiased manner

Redman and Parkerton (2015) identified characteristics of toxicity tests that make comparisons and interpretations among oils difficult. They highlighted how study design affects exposure to oil and the interpretation of toxicity, and presented a set of guiding principles to improve oil testing (Box 3). Principle 1 identified that the measured toxicity of an oil solution depends on the chemical concentration and composition of the oil solution, no matter which method was used to generate the test solution. Therefore, the characteristics of the oil should be considered in the experimental design. Principles 2, 3 and 4 noted that the concentration and composition of the dissolved phase in test solutions depends on the methods used to prepare test solutions, the proportion of particulate oil in the preparation, the headspace of the mixing vessel (particularly for the volatile components), and the composition of the test substance. Principle 5 stated that differences between exposure regimes (e.g., constant vs time-variable declining exposures) change the composition of test solutions and resulting toxicity. Based on toxicokinetic models, constant exposures allow more time for all components to reach equilibria between droplets and water and between water and tissues. In exposures with declining oil concentrations, the more soluble hydrocarbons are dominant in the beginning of the exposure but decline in importance over time. The more persistent PAHs are taken up less than in the constant exposure regimes but decrease to a lesser degree than the volatile components over time. It is important to note that the acute lethality of oil is attributed to the rapidly lost benzene, toluene, ethylbenzene and xylene (BTEX) and other volatiles in test solutions, and chronic toxicity is related to the more persistent PAHs. Principle 6 states that the dissolved phase can be characterized with passive samplers. From these principles, the authors developed a series of questions (Box 3) to evaluate the quality of eight recently-published toxicity studies and found that only two had at least partial answers to all of the criteria. Although Redman and Parkerton (2015) focused on acute lethality, their principles also apply to chronic toxicity tests, and provide a strong rationale for standard test methods that generate constant concentrations of hydrocarbons in water.

Box 3: Principles for oil toxicity testing

(Redman & Parkerton 2015)

Principles:

- The concentration and composition of dissolved hydrocarbons generally determines the aquatic toxicity of physically and chemically dispersed oils;
- Dissolved hydrocarbon exposures depend on WAF preparation method;
- Dissolved hydrocarbon exposures depend on amount of headspace in WAF test system;
- Dissolved hydrocarbon exposures depend on substance composition;
- Soluble hydrocarbons contribute more to toxicity under time variable exposures; and
- Passive samplers allow direct measurement of dissolved hydrocarbon exposures.

Evaluation Questions:

- Does the study report composition of hydrocarbon in the exposure media rather than nominal or measured total hydrocarbon concentrations?
- Does the study express different treatments in terms of dissolved phase toxic units to facilitate comparison of observed toxicity results between oils/dispersion methods?
- Is the dosing method (e.g., WAF, oiled gravel) and exposure chambers open or closed? If closed, is the head space volume reported?
- Is the substance composition reported to allow fate model predictions to be compared with observed exposure measurements for substance components?
- Are BTEX components considered?
- Have passive sampling measurements been used to allow measurement of dissolved hydrocarbon exposures?

Several recent publications have also critiqued specific papers that, in the opinions of the authors, have not been prepared, designed, conducted, or interpreted appropriately. For example, Coelho et al. (2013) and Redman and Parkerton (2015) reviewed Rico-Martínez et al. (2013) and identified key weaknesses in their test methods. These included a lack of standardized methodology (did not report the oil loading or mixing conditions) and a failure to measure total petroleum hydrocarbons (TPH) and TPAH to characterize exposures for comparisons of toxicity to other studies.

3.4.2 CROSERF protocol for oil toxicity tests (Aurand & Coelho 2005)

The most commonly used methods to mix oil and water for toxicity testing are the CROSERF protocols (Tables 5 & 6). They were developed for marine toxicity tests i.e., tests of marine species in natural or artificial salt water. The exposure methods employed depend on the objectives of the study and often on the anticipated nature of environmental exposures, e.g., whether the oil is floating on the surface of the water, dispersed into the water column, or interacting with sediments, and whether concentrations are constant or vary over time. Nevertheless, the goal of the exposure regime is to generate a series of exposure concentrations that cause a gradient of responses so that median lethal concentrations (LC50s), median effective concentrations (EC50s; sublethal responses) or threshold concentrations can be calculated.

Table 5 outlines the terminology for the three main solution preparation methods and six regimes for the exposure of aquatic organisms to the prepared test solutions. The exposure regimes vary in the stability of hydrocarbon concentrations over time. In constant exposures, the intent is to maintain constant concentrations over time, using static non-renewal (SNR; no replacement of test solutions over the course of the exposure), static renewal (SR; replacement of test solutions in a set time interval), or continuous flow (CFT) protocols. In contrast, spiked exposure (SE) protocols decrease hydrocarbon concentrations over time by continuous dilution of test solutions with fresh water, and in pulsed exposures, the test organisms are exposed alternately to fixed concentrations of test solutions and clean water, for fixed intervals. The differing methods, as well as the physical and chemical characteristics of the test oils, including dilbits, will affect concentration and composition of the test solutions (i.e., oil phase, droplet size distribution).

Table 5: Protocol Terminology from CROSERF (from Aurand and Coelho 2005).

Solution Preparation

PROTOCOL TERMINOLOGY	DESCRIPTION OF METHOD IN AURAND AND COELHO 2005	NOTES ON TERMINOLOGY USED IN THIS REPORT
Water Accommodated Fraction (WAF)	"… solution derived from low energy (no vortex when stirred by a magnetic stir-bar) mixing of test material which should be free of particulates of bulk material (>1 micron diameter)"	Corresponds to WSF and LEWAF from Table 2.
Physically-Enhanced WAF (PE-WAF)	"… high to very high energy (≥25% vortex) mixing of test material and water only (no chemical dispersant) in which a relatively stable population of bulk material droplets (1-70 micron diameter) is present"	Corresponds to MEWAF of Table 2. Some studies use WSF and WAF for the same procedure.
Chemically-Enhanced WAF (CE-WAF)	"… 20-25% vortex mixing of test material and chemical dispersant in which a relatively stable population of bulk material droplets (1-70 micron diameter) is present"	Corresponds to CEWAF of Table 2 (term used throughout this report). May also be called HECEWAF.

Exposure Regime

PROTOCOL TERMINOLOGY	DESCRIPTION OF METHOD IN AURAND AND COELHO 2005	NOTES ON TERMINOLOGY USED IN THIS REPORT
Constant Exposure (CE)	“refers to a constant exposure in which the aim is a constant concentration. In this type of exposure, the organisms are placed in a chamber and exposed to a test solution for a certain time period. The exposure may be flow- through, static non-renewal, or static renewal (see definitions). The tests are typically carried out for 48 or 96 hours.”	The CROSERF text implies that static non-renewal and static renewal protocols will generate constant exposures, but this is not the case (Chapter 4.3).
Continuous Flow-Through (CFT)	“refers to a constant exposure in which the organisms are placed in full strength test solution in a flow-through test chamber. The pumps are turned on and the chambers are pumped with a constant concentration test solution (i.e., no dilution of the test solution occurs). When performing this type of exposure, care must be taken to assure that components of the test solution are not being lost through the tubing walls.”	
Pulsed Exposure (PE)	“refers to a square-wave concentration exposure. In this type of exposure, organisms are placed in a closed, static chamber with no head space above the test solution. The organisms are left in the solution for a certain time period, then removed and placed in clean water. There is not a specified chamber type in which this type of test must be performed, nor is there a defined time period for the exposure to test solution prior to transfer to clean water. The tests are typically carried out for 48 or 96 hours.”	
Spiked Exposure (SE)	“refers to a declining concentration exposure. In this type of exposure, the organisms are placed in full strength test solution in the flow-through test chambers. The pumps are turned on and the chambers are slowly diluted with clean water. The standard flow rate used for dilution in the chambers (designed by UCSC) is 2 mL/min., which relates to an equivalent half-life (for the test solution) of 1.67 hours The tests are typically carried out for 48 or 96 hours.”	Note that the declining concentration test at 2 mL/min is based on a 240 mL CROSERF Flow-Through Toxicity Test Chamber
Static Non-Renewal (SNR)	“refers to a constant exposure in which the organisms are placed in a closed, static chamber with no headspace above the test solution. The initial test solution is used for the duration of the exposure (i.e.,-no test solution renewal).”	SNR exposure regimes are characterized by a decline in the test solution concentration over the duration of the exposure. (Aurand & Coelho 2005)
Static Renewal (SR)	“Refers to a constant exposure in which the organisms are placed in a closed, static chamber with no headspace above the test solution. The test solution is renewed at regular time intervals (typically 24 hours) with fresh test solution made to the same concentration.”	SR exposure regimes are also characterized by daily declines in the test solution concentration over the duration of the exposure (i.e., a series of spikes). (Aurand & Coehlho 2005)

Appendix 1 outlines the components of CROSERF protocols for preparing and using WAF and CEWAF test solutions, including the preparation of the test substance and test solutions, toxicity testing and outcomes, chemical characterization of test substance and test solutions, and statistical analyses. Some of the details are outlined below.

3.4.2.1 Methods for preparing test solutions (Details summarized in Appendix 1)

Test oils should be stored in the dark at < 5ºC in tightly-sealed metal or glass containers with minimal headspace; freezing should be avoided due to the separation of waxes and asphaltenes. Oil preparation will vary according to the weathering state of the test oil. No preparation is needed for tests with fresh oil. Artificial weathering should be 6-24 hours, following ASTM Method D2892 (fractional distillation method up to cutpoint 400°C) or D86, and equivalent to weathering during an actual spill. Natural weathering is not recommended as it is not reproducible. The proposed guidance in this review (Table 7) recommends that maximum temperatures should not exceed 200°C (D2892) or 130°C (D86). The weathered oil must be characterized chemically using the same analytical methods as the parent oil, preferably by a commercial lab.

CROSERF protocols were developed solely for marine ecosystems and do not include guidelines for preparing solutions with fresh water. Diluent sea water preparation should be the same for both WAF and CEWAF: a minimum filtration of 0.45 microns; good aeration; sourced from local or reconstituted sea water (sea salt diluted in deionized water); and salinity should match the requirements of the test organism. Diluent water and oil should be mixed in a sealed vessel (usually a cylindrical carboy) of appropriate size to produce the volume of solution required for each concentration of the test. The suggested minimum and maximum sizes are 1.0 and 20 L, respectively, with 20-25% headspace. For both WAF and CEWAF, a known volume of oil should be added to diluent water in a mixing vessel (0.01 to 25 g/L loading) after mixing has begun; delivery mass should be calculated by weight difference. For CEWAF only, a known volume of dispersant (e.g., Corexit 9500 and/or Corexit 9527) should be added sequentially to diluent water following oil at a dispersant-to-oil ratio of 1:10. Solutions should be mixed by a magnetic stir plate with a stir bar whose size will vary according to the volume of the mixing vessel; there is no requirement to report stir-bar size. While CROSERF originally specified a low mixing energy for WAF to avoid forming a vortex (180 to 240 rpm), and a higher mixing energy for CEWAF to produce a vortex that is 20-25% of water depth, the committee later recommended equal mixing energies for both solutions to better represent environmental conditions during a spill. A 20-25% vortex was chosen as representative of field conditions because it did not produce physical dispersion or emulsification in 1 to 20 L carboys. Mixing should be carried out in the dark or in laboratory light in a controlled temperature environment, depending on the temperature appropriate to the test organism. Mixing time should be based on oil type and loading, but 18 to 24 hours was recommended, in part to coordinate with daily static renewal exposures, followed by 3 to 6 hours of settling for CEWAF only. Solutions should be tested within 24 hours to prevent loss of volatiles and instability of solution composition and concentration. Samples taken for chemical analysis should be acidified immediately and analyzed within 7 days for volatiles (BTEX) and 14 days for semi-volatiles.

3.4.2.2 Exposure Regime and Test Outcome (Details summarized in Appendix 1)

Following solution preparation, test animals (temperate species, following ASTM, OECD, EPA guidelines for handling and care, or equivalent) should be added to air-tight chambers: solutions are placed in the open chamber, animals are transferred with minimal addition of clean water, and chambers are sealed. For CFT and SE protocols, test solutions or clean water are added through a top port.

CROSERF outlined protocols for two different methods of exposure for 96 h acute toxicity tests: spiked exposure testing and constant exposure testing. Spiked exposures add undiluted test solution prior to introduction of animals, and then flush exposure chambers with clean water at 2 mL/min after the addition of test organisms to a 240 mL CROSERF Flow-Through Toxicity Test Chamber. The expected half-life of the initial concentration is approximately 2.5 h, with a 95% decline within 12.5 h (though the decline should be monitored by chemical analyses within the first 6 to 8 hours of exposure). Spiked exposures were designed to reflect the rapid spreading and dilution of an offshore spill. To model nearshore or restricted waters, many additional considerations are necessary, including oil volume, dilution potential of receiving water (i.e., to simulate slower dilutions with 5 to 10 h half-lives), energy regime at the time of spill, and hydrographic conditions.

Constant exposure tests add test solution to glass-covered containers (20% headspace) followed by addition of test organisms. Every 24 hours, 90% of the test solution is replaced with fresh solution. Mortality and other effects should be monitored daily, though it is specified that dead animals not be removed. Endpoints should include lethality (96 h LC50), and initial narcotic effects (moribundity) (0.5-1 h EC50). The emphasis is on acute lethality and there are no recommendations for chronic exposure test methods.

Statistical methods should be reported whenever presenting estimates for endpoints or other toxicological data. CROSERF recommends specific tests for statistical analysis. If data are normally distributed, a probit test should be used to estimate toxicological endpoints. For non-parametric distributions (i.e., if no control mortality), a Spearman-Karber or Trimmed Spearman-Karber test should be used to estimate toxicological endpoints. Significant differences should be determined by comparing 95% confidence intervals. If no partial kills occur during a test (i.e., only 0 or 100% mortality), it is recommended to use either the binomial method, or to re-run the experiment at a different range of concentrations.

3.4.2.3 Test organisms

CROSERF recommended several fish species for toxicity tests, including Topsmelt larvae (*Atherinops affinis*), inland silverside juveniles (*Menidia beryllina*), and red drum larvae (*Scianops ocellatus*). Comparisons of oil toxicity between recommended and regional test species showed similar results (Aurand & Coelho 2005). Animal care and handling should follow published protocols by ASTM, OECD, and US EPA.

3.4.2.4 Methods to Characterize Oil and Test Solutions (Details summarized in Appendix 1)

For both WAF and CEWAF, CROSERF recommended a complete characterization of the initial oil by gas chromatography-mass spectroscopy (GC-MS). Prior to toxicity testing, total petroleum hydrocarbons (TPH, C_{10} – C_{36}) and volatiles (C_6 – C_9) should be analyzed using GC with flame ionization detection (GC-FID) in both WAF and CEWAF solutions; PAH analysis is optional. The same chemical characterization should be repeated at the end of the experiment for static tests. For renewal tests, samples should be taken at least once prior to each renewal. For spiked tests, samples should also be taken at one or more times during the exposure. Dispersant controls are a series of dispersant-only concentrations (1 to 1000 mg/L) in the same sea water used in toxicity tests. The dilution series may be used to create a calibration curve for analysis by ultraviolet-visible spectrophotometer (UV-VIS) within two hours of sample collection. The absorption maxima determined for the dispersant should be used for all dispersant analyses of CEWAF solutions.

Performance criteria for chemical characterization are needed to ensure valid inter-laboratory comparisons of hydrocarbons. TPH (equivalent to Total Extractable Organic Carbon (TEOC),

C_{10} to C_{36}) should be measured by GC-FID using the common baseline integration technique; resolved hydrocarbons ($TPH_{resolved}$) should be reported corresponding to individual compounds and not the common baseline integration technique. A TPH analysis should include at least one surrogate standard (o-terphenyl with 70 to 120% recovery), an internal standard (can be lab-specific as long as it does not co-elute with the test oil), and the results are not corrected for recovery. Another method besides GC-FID may be used, provided the analytical report compares it to the GC-FID method.

Volatile hydrocarbons should be measured using GC-FID or GC-MS, depending on available equipment and provided the instrument can identify the minimum target analytes (Appendix 2, Table 2A). Experimenters may choose to exceed this list, though this must be reported to ensure inter-laboratory comparison of studies. An internal standard should be used (can be lab specific as long as it does not co-elute with the test oil). D-toluene is the recommended surrogate standard for GC-MS analysis of volatile components.

Compound-specific semi-volatile analysis is optional due to cost and time restraints, though a list of analytes has been suggested (Appendix 1) based on the list of standard semi-volatile compounds analyzed during post-oil spill damage assessments. Semi-volatile component analyses may be more relevant in chronic toxicity tests (i.e., non-CROSERF) as these compounds, specifically the 3- to 5-ringed PAH with alkyl side chains, have been identified as the most chronically toxic constituents of oil to the early life stages of fish (Hodson et al. 2007; Adams et al. 2014a; Bornstein et al. 2014). Analytical methods should be based on available equipment and chosen with target analytes in mind. Again, it is essential to report measured analytes to ensure comparability of test data among labs.

The development of these methods required pooled resources for a substantial dataset. The main findings demonstrated that WAF and CEWAF have similar toxicity based on measured concentrations of volatiles and semi-volatiles, spiked exposures are less toxic than constant exposures, and weathered oil is less toxic than unweathered oil (Aurand & Coelho 2005).

The CROSERF Committee debated several topics in the development of the methods including variable loading versus serial dilution and open or closed test chambers. The Committee concluded that neither variable loading nor serial dilution could be recommended over the other. Despite a risk of overestimating toxicity, closed chambers were recommended over open chambers, which gave less consistent results due to variable evaporation (Aurand & Coelho 2005). The Committee also concluded that the protocols could be modified to include different dilution rates and different exposure chambers than those developed by CROSERF (Aurand & Coelho 2005).

3.4.3 Recommended changes to standardized methods with rationale

Modifications to the protocols to fit objectives outside of the original CROSERF initiatives have been undertaken (e.g., Barron & Ka’aihue 2003), recognizing that the primary focus of CROSERF was to develop standardized protocols to decrease inter-laboratory differences in dispersant and dispersed oil toxicity test results.

For example, there have been serious debates about the methods to create a gradient of oil exposures. CROSERF recommended the use of a variable loading gradient created by preparing solutions with a series of decreasing oil:water ratios. In contrast, Barron and Ka’aihue (2003) advocated variable dilutions, i.e., concentration gradients created by dilutions of a stock of test solution prepared with a single oil:water ratio. The two methods yield different exposures and toxicities. Variable loadings provide oil-to-water ratios that may mimic conditions of a spill on the open sea. However, there are differences among treatments in the hydrocarbon composition of the aqueous phase of oil due to the different solubilities of hydrocarbons. Also, variable loadings do not model dilution following the mixing of a surface slick into the water column. These differences complicate the interpretation and comparisons of measured toxicity among different oils. Variable dilutions promote consistent chemical composition and comparability among test solutions of dispersed oil, and are simpler to prepare.

Barron and Ka'aihue (2003) also recommended changes to the CROSERF protocols to facilitate dispersant tests with sub-Arctic species, under sub-Arctic conditions, including lower temperatures and longer exposure periods (Summary, Box 4). These modifications were justified based on environmental differences in sub-Arctic conditions (i.e., UV light, temperature) and potential spill conditions (i.e., 1:10 DOR). The method of diluting WAF to low concentrations in the CROSERF flow-through method may underestimate the persistence of oil in Arctic environments. Conversely, in sub-arctic environments, the prescribed sealed test vessels may increase the exposure to monoaromatic hydrocarbons, which are likely lost within the first 48 hours of a spill at higher temperatures (Barron & Ka'aihue 2003).

Box 4: Changes recommended to CROSERF protocols for dispersed oil toxicity testing under Arctic conditions

(Barron & Ka'aihue 2003)

- Test both fresh and moderately weathered oil under conditions of moderate mixing energy
- Test both Corexit 9500 and 9527 using high dispersant:oil ratios
- Prepare test solutions by variable dilutions rather than by variable loading
- Use static exposures in open chambers
- Increase the test duration from 4 to 7 days to assess delayed mortality
- Quantify approximately 40 polycyclic aromatic hydrocarbons (PAHs) and their alkyl homologs in toxicity test solutions
- Test Pacific herring larvae, an urchin species, and a calanoid copepod under sub-Arctic conditions of temperature and salinity
- Assess the potential for photoenhanced toxicity by incorporating limited exposures to sunlight or simulated natural sunlight
- Incorporate a bioaccumulation endpoint by measuring PAH concentrations in copepod tissues.

Barron and Ka'aihue (2003) also recommended the addition of a UV light regime with a 24-hour delay from initiation of the oil exposure to test the potential exposure to phototoxic components. Additionally, further clarification of the loading of oil in water could entail reporting the thickness of the floating oil layer during preparation of stock solutions. Many of these recommended changes improved the protocol by making it more practical (e.g., using open chambers) and shifting the emphasis from 'realistic' to 'worst case' scenarios (e.g., photo-enhanced toxicity; high oil:water ratios).

NRC (2005) recommended changes to the CROSERF protocol which included:

- The preparation of WAF and CEWAF with equivalent mixing energies;
- Flexibility in the choice of different test chambers as long as they are consistent for all exposures;
- Additional guidance on requirements for chemical analysis of test solutions and sampling regimes, calculation of endpoints, and calculation of toxic units;

- Additional methods to address UV photo-enhanced toxicity.

Clark et al. (2001) recommended equal mixing energies for preparing WAF and CEWAF, noting that although there will be more variability in the WAF stock solutions with particulate oil, that this scenario is more realistic to a spill. NRC (2005) also recommended that the protocol for declining oil concentrations be modified to include a slower rate of decline; Aurand and Coelho (2005) identified this amendment as the most useful addition to the CROSERF protocols at the time of their report.

A review of publications that cite the CROSERF guidance document demonstrates a number of modifications to the basic method to make it more efficient and pragmatic, or to suit the specific needs of research (Chapter 4 below). For example, Adams et al. (2014a) describe a high energy CEWAF (HECEWAF) method that applies high energy mixing of water, oil, and dispersant with a 100% vortex followed by sonication. The purpose was to reduce the influence of the high viscosity of heavy or highly-weathered oils on their dispersibility and bioavailability. The method generated stock solutions rich in dissolved and particulate oil and allowed comparisons among oils based on the inherent toxicity of their constituents, not toxicity modified by physical limitations to solubility.

4 LITERATURE REVIEW

Chapter 3 outlined the CROSERF guidance document with methods for testing the toxicity of dispersed and chemically-dispersed crude oil in sea water. To date, standardized methods have not been developed for alternative scenarios such as freshwater spills, or unusual products such as heavy oil and dilbit. Chapter 4 reviews the potential effects on oil toxicity tests of different protocols reported in the literature related to oil storage, preparation of oil solutions, exposures of test species, chemical analyses, etc. The review supports recommendations in Chapter 5 for modifications to the CROSERF guidance in a proposed framework and for research to address critical unknowns.

4.1 METHODS FOR LITERATURE REVIEW AND ANALYSIS

This review included database searches with predetermined search criteria focusing on peer-reviewed reports of oil toxicity tests and test methods with fish. Databases reviewed included Google Scholar, Web of Science, and Queen's University Library Summons (key words in Box 5). The scope of this review was limited to laboratory methods for oil toxicity tests with fish. These studies are numerous and provide substantial information about advantages and disadvantages of available methods, knowledge gaps, and areas for future research. The review was further restricted to tests in which oil was mixed with water, or its components were added by partition-controlled delivery systems with oiled substrate. Oil-contaminated sediments can also be a source of waterborne hydrocarbons and a direct source of exposure for species that filter or consume sediments (e.g., gizzard shad (*Dorosoma cepidianum*)). Although exposure of benthic species to heavy oil and dilbit associated with

Box 5: List of key words in the literature review

- PAH toxicity oil dispersant
- dispersant Corexit chemical toxicity
- dispersant Corexit toxicity
- Corexit oil toxic fish
- oil dispersant toxicity in fish
- fish toxicity crude oil
- CEWAF toxicity fish
- PAH oil toxicity tests
- oil toxicity
- toxicity dispersant crude oil fish
- oil toxicity test Corexit
- AMOP CEWAF
- IOSC CEWAF
- SETAC CEWAF
- IOSC CEWAF methods review

sediment is a major concern due to their propensity to sink, sediment toxicity test methods were not included in this report. The review also excluded tests of oil-contaminated water from experimental or actual oil spills.

The components of toxicity test methods were separated under four categories: methods for preparing test solutions; methods to characterize oil and test solutions; exposure regime; and test outcome. The descriptive statistics and analyses for each of these four categories are presented below. If a publication did not report a component, it was categorized as **NS,** information not specified. Components that were not applicable to the methods in the publication were designated as not applicable, **N/A**.

The compilation of data was limited to the components of protocols that were explicitly reported in each publication. Methods cited from other publications were included only if they were described in sufficient detail to ensure that the methods were followed as described. This approach may have increased the number of **NS** entries.

The summary of the literature captured is in Box 6. The numbers in each category do not add up because multiple tests (e.g., oils, species) were included in many papers. Most papers were published within the last six years (Figure 2) which likely reflects the interest and funding for research following large volume oil spills, including the Deepwater Horizon well blowout. The oldest paper included in the review was published in 1973 and the most recent in 2016.

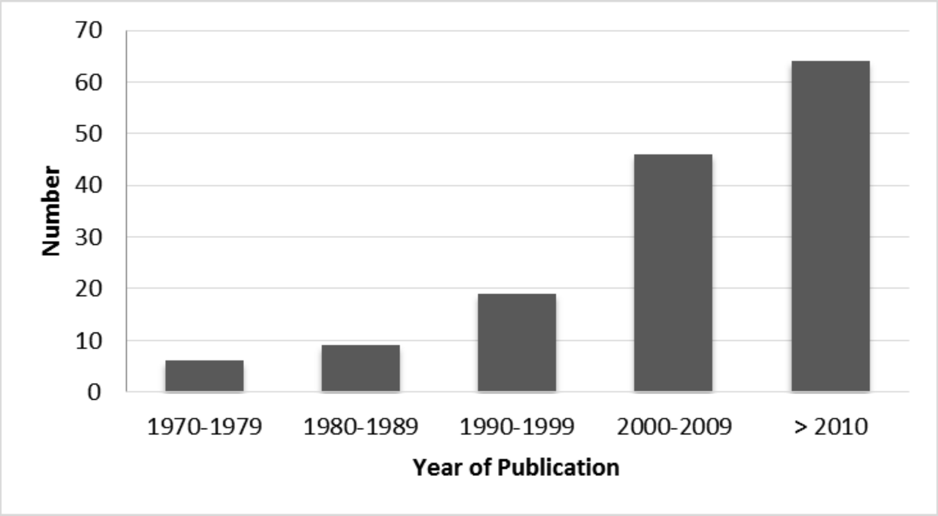

Figure 2: The number of papers included in the literature review by year of publication. Range of years between 1973 and 2016 (N = 144).

4.1.1 Identification of experimental factors that affect the outcome of toxicity tests

The review mainly covers toxicity tests using WAF and CEWAF exposures, as these are the focus of CROSERF. The concentration and composition of WAF and CEWAF are affected primarily by the oil loading to water, the dispersant-to-oil ratio (DOR), temperature, and the energy and duration of mixing. Dispersants make oil more bioavailable by increasing the surface-to-volume ratio of oil droplets, increasing their stability in the aqueous phase, and maximizing the partitioning of hydrocarbons from oil to water (e.g., Ramachandran et al. 2004; Schein et al. 2009; Wu et al. 2012; Martin et al. 2014). Heavy oils may appear less toxic than lighter oils with similar concentrations of toxic components because the higher viscosity of heavy oils limits the formation of small droplets under the low mixing energies typical of WAF preparation.

To assess the impact of experimental factors on the test outcome, all factors need to remain unchanged except for the one variable under assessment, but few studies have systematically assessed the effects of each factor on test results. Echols et al. (2016b) compared the acute lethality of three different oils to juvenile inland silversides (Menidia beryllina) and mysid shrimp (Americamysis bahia) to assess the influence of four variables prescribed in the CROSERF method (Singer et al. 2001): open versus closed static test chambers, natural versus synthetic water diluents, aerated versus non-aerated water, and low (LEWAF; no vortex) and medium mixing energy (MEWAF; 25% vortex) WAFs. The test oils included one unweathered and two weathered samples from the Deepwater Horizon oil spill. Overall, unweathered oil was more acutely lethal than weathered oil, corresponding to higher concentrations of low molecular weight hydrocarbons in test solutions. Shrimp were more sensitive to oil exposure than fish, but many tests with fish were inconclusive because 96 h LC50s exceeded 100% WAF in most tests of weathered and unweathered oil. The composition of LEWAF did not change when solutions were prepared with natural sea water, carbon and micro-filtered sea water, or synthetic sea water with Instant Ocean. Although the nature of the dilution water did not change the toxicity of weathered or unweathered oil to fish, shrimp were more sensitive to oil in artificial and filtered salt water than in natural salt water; the cause of this difference was not explained. The toxicity of unweathered oil increased by about 50% with increased mixing energy, consistent with higher concentrations of hydrocarbons in MEWAF than in LEWAF. As well, the chemical composition and toxicity of LEWAF to shrimp were more reproducible than WAFs prepared with higher energy mixing, likely due to lower concentrations of particulate oil. Compared to closed test vessels, there was a loss of C1 and C2 naphthalenes from solutions of unweathered oil in open test vessels, indicating weathering and corresponding to a reduction in toxicity to shrimp. Gentle aeration of test solutions had no effect on toxicity. While these experiments provided guidance on test methods, particularly with shrimp, conclusions about the effect of test variables on toxicity to fish were weakened by the low acute lethality of the oils tested and the absence of statistical tests of treatment effects. When 96 h LC50s exceeded 100% dilutions of WAF, there was no basis for assessing treatment effects.

Rhonton et al. (2001) compared the acute toxicity of chemically-dispersed unweathered and weathered ANSC to larvae of estuarine fish (*Menidia beryllina*) in 96-hour spiked and continuous exposures. Based on nominal oil-loadings, CEWAF was more toxic than WAF and continuous exposures were more toxic than spiked exposures. Nevertheless, when 96 h LC50s were expressed as measured TPH, the 96 h LC50s for continuous exposure to WAF and CEWAF were similar for both weathered states of ANSC. Similarly, the acute toxicity of WAF of unweathered and artificially weathered PBCO was lower than that of CEWAF when based on nominal loadings, but equitoxic when based on measured concentrations of hydrocarbons (Singer et al. 2001b). Redman and Parkerton (2015) reviewed methods for oil toxicity testing and computing LC50s of toxic units derived from chemical analysis of test solutions and the PETROTOX model. The toxicity of the WAFs prepared by different methods overlapped when expressed as measured concentrations and toxic units (TUs) of hydrocarbons, particularly when test solutions were characterized by passive sampling devices that enhanced detection limits. Similarly, for chronic toxicity, CEWAFs of four crude oils (Wu et al. 2012), diesel oil (Schein et al. 2009), and heavy fuel oil (Adams et al. 2014b) were more toxic to trout embryos than corresponding WAFs when expressed as percent dilution (nominal concentration), but virtually the same when expressed as measured concentrations of TPAH or total petroleum hydrocarbons measured by fluorescence (TPH-F) in solution.

In contrast, the chronic toxicity of CEWAF prepared from heavy fuel oil was lower than WAF generated from an oiled-gravel column, even when toxicity was expressed as measured concentrations of TPH-F (Martin et al. 2014). The apparent reason was differences in the composition of test solutions. Hydrocarbons in CEWAF included both particulate and dissolved oil, with a composition similar to that of bulk oil. In contrast, WAFs from oiled columns were predominantly dissolved hydrocarbons, with a different distribution of PAH from the parent oil. This illustrates a common bias in reports of the concentrations of oil causing toxicity because the extraction of water samples with solvent incorporates both the particulate and dissolved phases of oil in test solutions. As a result, the toxicity of the dissolved components of oil solutions (e.g., BTEX for acute lethality, PAH for chronic toxicity) may be underestimated. This reinforces the recommendation by Redman and Parkerton (2015) to use passive sampling devices that target the dissolved component of oil test solutions.

Tsvetnenko and Evans (2002) assessed how mixing energy, mixing duration, and volume-to-interface ratio of floating oil affected the composition of WSF prepared from light (42 ^{o}API and 48 ^{o}API), medium (34 ^{o}API), and heavy (21 ^{o}API) crude oils and a natural gas condensate (44 ^{o}API), and the associated toxicity to unicellular algae. The goal was to identify aspects of standardized methods that influence the outcome of laboratory tests with unusual oils, specifically the highly volatile crudes from the North West shelf of Australia. The standard method was the preparation of WSF with mixing of oil in filtered and sterilized natural (35 ppt salinity) sea water for 20 hours, and exposure of unicellular algae to the aqueous phase for 96 hours. The authors found that:

- WSF generated in 1 L cylindrical bottles with a fixed 1:9 OWR, a lower total volume of oil and water, a thinner layer of oil on water, and an increased headspace reached the maximal hydrocarbon concentration 3-fold faster than preparations with larger volumes of oil and water;
- Increasing stir bar spinning from 140 to 540 rpm increased the depth of the vortex for all stir bars tested (25, 34, 38, 49 mm length, wedge 51 mm, findisk 35 mm). The recommended stir bar size for a stable vortex was 34 mm at 160-190 rpm. The energy applied to attain a given depth of vortex depended on the size of the stir bar;
- Concentrations of heavy oil in WSF increased continually during 600 hours of mixing, whereas other oils reached a maximum loading within 24 hours. Increasing the temperature of mixing decreased the time to maximal concentration of the WSF, but temperatures should not exceed that of the toxicity test;
- The toxicity of WSF of the lightest oil did not change after 24 to 240 hours of continuous stirring despite differences in chemical composition due to volatilization of hydrocarbons prior to and during the toxicity tests; ultimately there were similar concentrations of hydrocarbons and similar toxicities.

Static exposures are more cost-effective, but create test solutions that change over 24 hours if implemented with a daily renewal regime (Echols et al. 2016b). Flow-through air-tight vessels increase the volatile aromatic fraction (along with other volatiles) in the WAF, which may overestimate the toxicity of this fraction in real spills where weathering rapidly reduces concentrations. Daily static renewal may also overestimate toxicity because daily dosing does not replicate typical marine spill scenarios. However, the Deepwater Horizon well blow-out and the pipeline spill of dilbit to the Kalamazoo River demonstrated that there is no single ‘spill scenario’. As with all spills, conditions vary widely according to the source and nature of the oil, the rate and duration of discharge, and the environmental conditions that control oil fate, distribution and exposure of receptors (Lee et al. 2015).

In general, there was not a lot of focus on water quality factors (pH, temperature, dissolved oxygen) for the tests. Readers must assume that these parameters were optimized and tracked, because they can modify test results by changing the fate and behaviour of oil and the responses of test organisms.

Research Needs

- Assess the fate, behaviour and effects of dilbit in fresh water, sea water, and brackish water, and under temperate and Arctic conditions.
- Compare the behaviour and toxicity of dilbit among marine, brackish and fresh waters, including those that are rich in humic acids or very low in alkalinity (also hardness, pH, other parameters).
- Test light and heavy crude oils at multiple weathering states to determine how chemical composition and toxicity of test solutions changes with mixing times and temperatures at different times during a spill event.
- To create perspective on relative risk, compare the toxicity of dilbit to that of previously-tested conventional crude oils, using the same standard methods for preparing test solutions and measuring toxicity.
- Develop standard methods to assess the toxicity of dilbit incorporated into sediments.
- Apply effects driven chemical fractionation (EDCF) to identify the toxic components of oil.
- Develop species sensitivity distributions based on standard oil toxicity tests.
- Assess the contributions to toxicity of products of photodegradation and biodegradation products and acid-extractable organics.

It is likely that test conditions also affect the outcome of dilbit toxicity tests, but to date there have been no systematic studies.

4.1.2 Oils tested

Weathering of dilbit and other oils during preparation of stock and test solutions has important implications for oil dispersion and the composition of WAF and CEWAF. The API gravity of dilbit decreases considerably with weathering, a process that dramatically changes its behaviour, fate, and chemical composition (Chapter 2). For instance, the API gravity of CLB dilbit is 22.6 ^{o}API so it is classified as a medium oil, just above the heavy-oil cut-off of 22.3 ^{o}API. The API gravity of undiluted Cold Lake bitumen ranges from 9.8 ^{o}API to 13.2 ^{o}API (EC 2016), suggesting that, when CLB dilbit weathers and the diluent volatiles are lost, CLB will become a heavy oil with a lower API gravity. Few papers included these details in the description of the oils used. *API gravity should be reported for all oils tested, if available, or calculated from the specific gravity.*

All 144 papers reported the study of at least one oil (criteria for inclusion), although some studied multiple oils, capturing 75 different identified oils (Figure 3). The most commonly tested oil was Alaska North Slope crude (**ANSC**), reflecting the intense interest in the 1989 *Exxon Valdez* spill in Prince William Sound, as well as the availability of oil for testing. Other commonly tested oils include diesel (**DSL**), Arabian Light crude (**ALC**), Prudhoe Bay crude oil (**PBCO**), Number 2 fuel oil (**N2FO**), Medium South American crude (**MESA**), Southern Louisiana sweet crude (**SLS**), and Macondo crude (**MAC**). The number of studies of MAC is increasing rapidly as research following the 2010 Deepwater Horizon spill is published. More than 50% of oils were studied four or fewer times and were classified as **Other**. When grouped by type (e.g., crude oil, heavy fuel oil, dilbit) the most frequently tested oil was crude (>70%), followed by diesel and heavy fuel oils.

In this review, the most commonly tested oils were also classified as light, medium, or heavy by their API gravity values. Most reports did not give an API gravity value, so in these cases the oil was classified by cross-referencing to API gravity values from the Oil Properties Database maintained by the Environmental Technology Centre at Environment Canada (2016), from MARPOL (2005), or from Dietrich et al. (2014). It is important to note that API gravity is not a fixed value for each oil, but varies with weathering, temperature, and the database or source, so that some oils at the margins may fall into more than one category.

Five of the eight most commonly studied oils (DSL, MAC, SLS, N2FO and ALC; Figure 3) were classified as light and the remaining three (ANSC, MESA**,** and PBCO) as medium. The abundance of tests of light oils reflects their high value and the large quantities shipped. As the production of light oils declines in the future, there will likely be a greater reliance on medium and heavy oils such as dilbit, higher volumes shipped, and a greater emphasis on toxicity testing and ERAs, particularly for freshwater ecosystems in Canada.

Only 22 of 144 papers compared oil toxicity and chemical characteristics among multiple types of oils; fuel oils (**N2FO***,* **HFO**, etc.) and **ANSC**, **ALC**, and **MESA** crudes appeared multiple times. Fifteen of these studies compared toxicity among light, medium, and heavy oils. This includes a comparison of heavier, high viscosity oils (MESA, Terra Nova) to lighter and less viscous crude oil (Scotian Light) (Ramachandran et al. 2004; Wu et al. 2012), comparison of the chronic toxicity of WAFs and CEWAFs of both MESA and HFO 7102 to embryos of Atlantic herring and rainbow trout respectively Adams et al. (2014b), and a comparison of dilbit to medium sour composite and mixed sweet blend crude oil (Philibert et al. 2016).

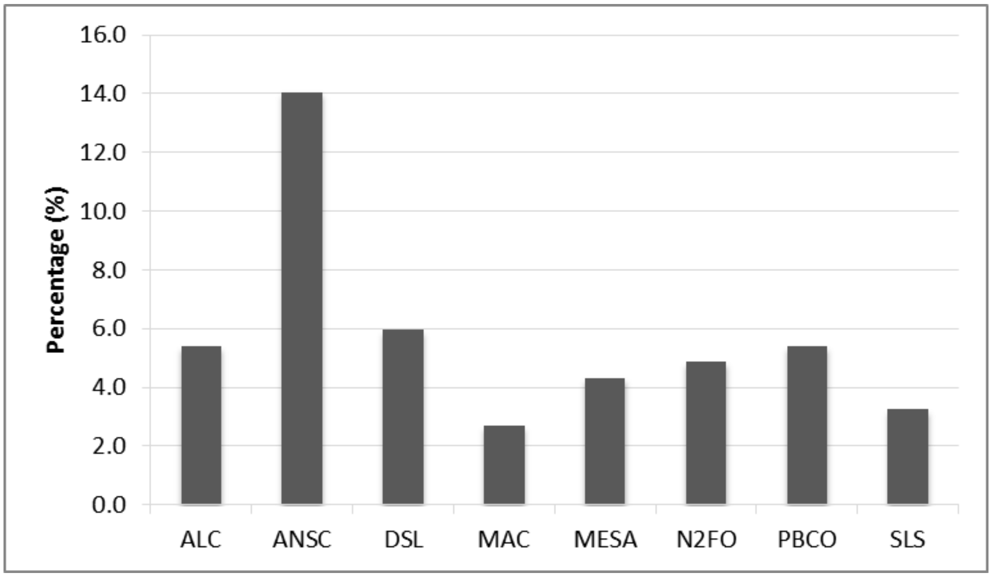

*Figure 3: The percentage of toxicity tests with different crude oils. Alaska North Slope crude oil (**ANSC**), Arabian Light crude oil (**ALC**), diesel (**DSL**), Prudhoe Bay crude oil (**PBCO**), Number 2 fuel oil (**N2FO**), Medium South American crude oil (**MESA**), heavy fuel oil (**HFO**), Macondo crude oil (**MAC**), Southern Louisiana sweet crude oil (**SLS**), and North Sea crude oil (**NSC**) (N=185 oil entries). Oils classified as **Other** are those that occurred four or fewer times among publications, and were not included in the graphic (N=100).*

4.1.3 State of weathering of oils tested

Weathering of oil results in changes to chemical composition as a result of processes including volatilization, dissolution of water soluble components, biodegradation, uptake of solid matter and/or water, emulsification, and/or photo-oxidation (Dew et al. 2015). It is likely that important weathering processes occur during all stages of toxicity testing, from receiving the oil to the final tests in the lab. This is especially relevant for dilbit because it rapidly loses the low molecular weight components of diluent due to volatilization (Chapter 2).

Unfortunately, detailed compositions of tested oils are rarely reported, and the terms ‘unweathered’ or ‘weathered’ do not accurately represent the state of a test oil. Weathering represents a continuum of changes in composition that occurs during the storage, handling, preparation and testing of oil solutions. The most appropriate terms for oils should be ‘less weathered’ and ‘more weathered’, but insufficient analytical data are typically reported to support these classifications. For the three reports of dilbit toxicity reviewed in Chapter 2, dilbit was classified as ‘unweathered’, but it is likely that weathering occurred during storage, handling, and testing.

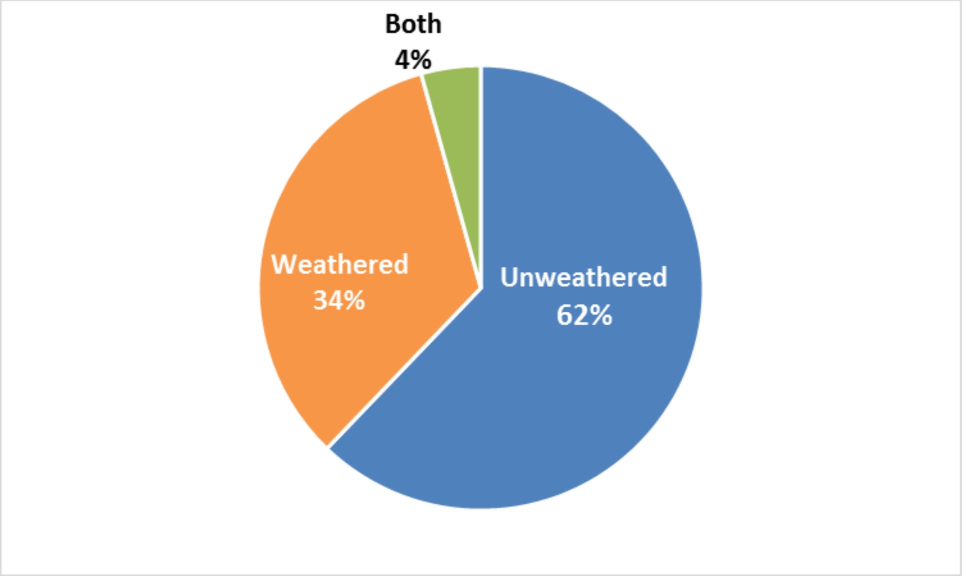

*Figure 4: The percentage of publications that used oils that either did not undergo weathering (**Unweathered**) or did undergo weathering (**Weathered**). If no weathering state was specified in the text, the oil was assumed to be unweathered. If one oil was tested at multiple weathering states, that oil was classified as **Both** unweathered and weathered (N = 185 oil entries).*

The 185 oils tested across all studies were classified according to their state when tested as **Unweathered, Weathered** or **Both** (Figure 4). If multiple oils were tested in a single study, each was given an individual classification. If the state of the oil was not described explicitly, it was assumed to be unweathered, which created a bias towards the number of unweathered oils. A weathered classification was given if the oil was weathered or was handled in a way that would result in weathering during the experiment. Most oils tested (62.2%) were unweathered and 33.5% were weathered.

A small number of reports compare weathered and unweathered versions of the same oil to specifically examine weathering. Gonzalez-Doncel et al. (2008) investigated the effects of unweathered and weathered Prestige crude oil on embryo-larval development of Japanese medaka, and found that while exposure to both states induced developmental abnormalities, TPAH was higher in the unweathered oil than the weathered oil and corresponded with higher toxicity. *Toxicity test reports should include chemical analyses of the test oil, preferably at the start and completion of a program of toxicity testing.*

Although CROSERF recommended that artificial weathering of oil should follow ASTM Method D2892 or D86, temperatures above 130˚C must not be used to avoid chemically changing the oil, including dehydration of cycloalkanes and breaking of chemical bonds. Under vacuum distillation conditions (ASTM Method D1160), the maximum recommended temperature is

200˚C (H. Dettman, Natural Resources Canada, personal communication). There are many *ad hoc* methods reported, including the ANS 521 weathered crude oil prepared by heating ANSC to 521˚F (272˚C, described in Bidwell et al. 2003), a technique that may degrade hydrocarbons and alter chemical composition and toxicity. The origin, state of weathering, and detailed methods for experimental weathering of test oils must be clearly reported in studies of weathered oil toxicity to ensure that results are comparable among studies and useful in ERAs.

Research Needs

- Assess the effects of oil weathering during solution preparation on the density (API Gravity) and viscosity of the test oil, on the chemical composition of test solutions prepared by standard methods, and on the measured toxicity of oil, to identify potential bias in test results and the implications of weathering on environmental fate and effects at actual spills.
- Characterize the nature and extent of physical and chemical changes to test oils under different storage conditions, and the potential effects on the outcome of toxicity tests.

Even when studies use unweathered oil, it is likely that these products are weathered to some extent during handling, preparation of test solutions, and storage due to the rapid loss of low molecular weight components. Weathering will alter the viscosity of the oil, dispersion in water, partitioning of hydrocarbons to water, and measurements of toxicity. *Therefore, it is critical to assess weathering by measuring weight and physical and chemical characteristics at the beginning and end of experiments.* Research is needed to identify what metrics indicate weathering during testing and how weathering might affect toxicity.

4.2 METHODS FOR PREPARING TEST SOLUTIONS

4.2.1 Frequently cited methods for preparing test solutions

The methods cited for preparing test solutions were captured for each publication to identify the most influential. Only a few papers (<13%) did not cite other methods in their description of the preparation of test solutions, and the number of citations of other methods varied from 1 to >10.

The most frequently cited methods were Singer et al. (2000) and Anderson et al. (1974), likely because these methods are associated with the standardization process. Many papers reported methods that included elements from both sources. Publications were classified as having cited the CROSERF method if they mentioned CROSERF directly, including the final report by Aurand and Coelho (2005), or indirectly if the publication cited a modified CROSERF method, such as Singer et al. (2000). Of 144 publications reviewed, 30 (20.8%) cited the CROSERF method, seven directly and 23 indirectly. Although many publications cited the CROSERF methods, most did not report all of the specific details or justify any reported modifications, and it can only be assumed that the method was actually followed as prescribed. It is critically important that all methods be described in detail, particularly modifications to cited protocols. Reports should also include all relevant characteristics of the test, from the nature of the test organisms to the measured concentrations of oil in water (see Table 7 and Appendix 4 for guidance on reporting). The development and use of standard methods also does not ensure that the tests produce high quality data. The assessment of the quality of data is outside the scope of this report. There are guidance documents that are helpful for assessment of ecotoxicology studies (e.g., Harris et al. 2014) and for oil toxicity tests (e.g., Redman & Parkerton 2015).

Few papers have described modifications to the CROSERF methods (Chapter 3) in sufficient detail that the modified methods can be easily repeated (e.g., Barron & Ka'aihue 2003; Hemmer 2011). An exception is Gardiner et al. (2013) who described how the protocol was modified to test oil toxicity to Arctic species under Arctic conditions (see section 4.2.2). They listed the modifications and provided descriptions of each component of the method, including the species tested, temperature, OWR and DOR, spiked exposure, extended test duration, and frequency of chemical analyses of test solutions.

There are several other methods that are quite different from the CROSERF protocols, particularly emerging methods for continuous-flow tests, including WAFs prepared by water upwelling through oiled substrates (Marty et al. 1997) and introduction of oil to flowing water by high pressure jets to create suspensions of dispersed oil (Nordtug et al. 2011). As with static toxicity tests, many of the method details that vary among publications have not been systematically assessed to determine how they affect the chemistry and toxicity of test solutions. Additionally, within research groups, methods can evolve over time without systematic experimentation to test the effects or published details of the changes made.

Methods used to prepare WAF and CEWAF stock solutions of dilbit were modified from Singer et al. (2000) and WAFs were also prepared by water upwelling through oil-coated ceramic beads (Kennedy & Farrell 2005). Although no specific modifications for preparing stock and test solutions with dilbit were noted in these studies, the resistance of dilbit to chemical dispersion and the potential for weathering during test preparation was identified as potentially contributing to test outcomes (Madison et al. 2015). *It is recommended that studies of dilbit include treatments prepared with standard methods as controls to account for the physical characteristics of dilbit that could modify test results (see section 4.2.2, bottom).*

4.2.2 Mixing instrument or system to generate stock and test solutions

The toxicity of oil in water depends on the concentrations of hydrocarbons in solution, which depend in turn on the amount, size distribution and stability of suspended oil droplets. The importance of the mixing method is the extent to which it promotes the partitioning of hydrocarbons from the oil phase to the dissolved phase, where they are bioavailable to test fish. Mixing methods that provide higher concentrations of small droplets that are stable in solution generate higher surface-to-volume ratios for partitioning and a longer period before droplets are removed from suspension (e.g., by coalescence). As a result, toxicity is more closely related to the measured concentrations of hydrocarbons in test solutions, not the nominal concentrations of dispersed oil added to a test solution (Schein et al. 2009; Wu et al. 2012; Gardiner et al. 2013; Adams et al. 2014b; Martin et al. 2014). For example, Gardiner et al. (2013) compared the toxicity of ANSC WAF, BWWAF (Breaking Wave WAF - prepared by intermittent manual shaking), and CEWAF to Arctic cod and sculpin. The three methods generated solutions with different amounts of oil droplets. Compared to stock solutions of WAF, measured concentrations of TPH and TPAH were 3.3- and 2.8- fold greater in BWWAF, and 194- and 119-fold greater in CEWAF. The composition of stock solutions also changed as alkylated PAH increased from 50% of TPAH (WAF) to 62% (BWWAF) and 87% (CEWAF). Compared to WAF and BWWAF, the proportion of alkyl PAH in CEWAF was closer to the parent oil, indicating that the CEWAF stock solution was dominated by free-phase oil droplets, with a much smaller proportion of dissolved hydrocarbons. Despite the change in composition and concentration of hydrocarbons in WAF, BWWAF and CEWAF, their acute toxicity was similar when expressed as measured naphthalene concentrations. Assuming that most naphthalenes are dissolved, this suggests that fish responded to dissolved rather than particulate oil, and that a greater dispersion of oil droplets increased the exposure to dissolved hydrocarbons. To understand the interactions between methods for preparing solutions and measured toxicity, it is essential that hydrocarbon concentrations be measured in test solutions.

The mixing system includes the equipment and/or method used to prepare oil/water stock and test solutions in the toxicity tests included in this review. A total of 226 mixing methods to generate solutions were recorded in the 144 publications reviewed, including 20 unique instruments or oil-water mixing systems (Figure 5). Almost half of the publications used more than one mixing system, which explains why the number of mixing methods exceeds the number of publications. The most common mixing instrument or method was the magnetic stirrer (**MS**) method specified by CROSERF (Aurand & Coelho 2005), the modified CROSERF method (Singer et al. 2000), and Anderson et al. (1974). Stir bar size determines vortex depth and stability and the extent to which dissolved and particulate hydrocarbons are incorporated into stock solutions. Although CROSERF guidelines (Aurand & Coelho 2005) recommend a 2-inch (50.8 mm) stir bar for 2 L of solution and a 1-inch (25.4 mm) stir bar for 1 L of solution, only Tsvetnenko and Evans (2002) have systematically defined the optimal stir bar size (37 mm) and shape for their test vessels (1.0 L conical flask) for generating uniform stock solutions. Few papers report the stir bar sizes or shapes used in stock solution preparation or other stirring conditions.

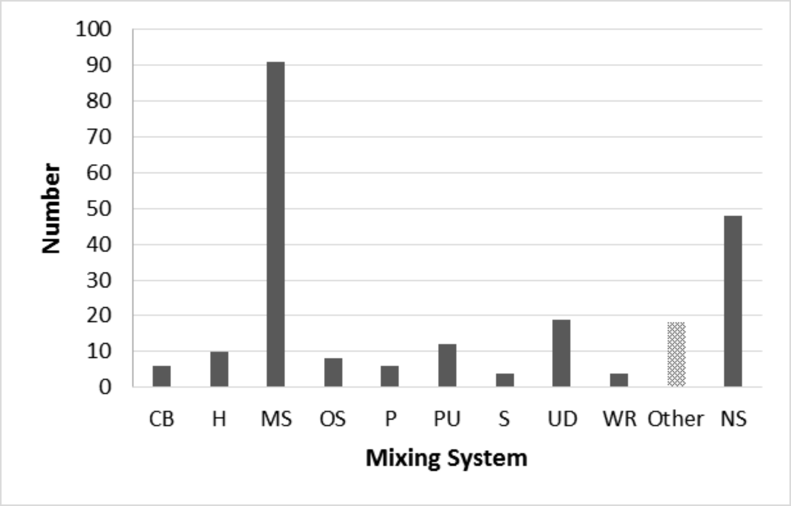

*Figure 5: The number of toxicity tests that used a particular mixing system to prepare stock and test solutions of oil in water. Some studies used more than one mixing system in a single experiment (N=226). Commercial blender (**CB**); Hand mixing (**H**); Magnetic stirring (**MS**); Orbital shaker (**OS**); Propeller (**P**); Pump (**PU**); Sonicator (**S**); Upwelling dilution (**UD**); Water recirculation (**WR**); Mixing systems occurring three or fewer times in publications (**Other**); Mixing system Not Specified explicitly in the publication (**NS**). One paper was omitted as there was no mixing of oil and water.*

The second most common mixing system was upwelling dilution (**UD**). A continuous flow of water upwelling through oiled substrates (e.g., gravel) created test solutions for toxicity testing by partitioning of hydrocarbons from a stranded oil phase to a mobile aqueous phase. The next most common systems were pumps (**P**) to mix and circulate oil and water and hand mixing (**H**). Many publications (21.2%) were classified as Not Specified (**NS**) because preparation of solutions was not adequately described. For example, preparation was described simply as “stirred slowly” (Ernst & Neff 1977), “stirred” (da Cruz et al. 2012), and “mixed” (Shuka et al. 2007). Other mixing systems included: flow-through toxicant apparatus, glass rod, hand vortex, “mixer”, motor paddle shaft, plastic circle, paint mixer, and top stirring. This lack of detail prevents the replication of experiments and comparisons of results among studies.

The method used to generate test solutions usually reflected the purpose of the study (Table 5 and Figure 5). For example, magnetic stirring was used frequently to create WSF, WAF, or CEWAF solutions to simulate episodic oil spills, while upwelling dilution was common in tests to simulate the embryo toxicity of contaminated spawning shoals. Mixing systems also depended on whether a single concentration of test solution was being prepared (CROSERF method) or a stock solution that was diluted in static or semi-static systems or in a continuous flow or constant exposure regime. Magnetic stirring was most frequently used in static or semi-static applications, whereas upwelling dilutions, propellers, and pumps were more frequently used in continuous-flow systems.

Other unique systems used to mix oil and water included a mixing tank partitioned into four compartments to mix oil and water while preventing small oil slicks and suspended droplets from reaching the adjacent exposure tank (Danion et al. 2011). A pumping system connected to the mixing and exposure tanks recirculated water between the two. A continuous flow of the WSF of Cook Inlet crude oil was prepared by dripping water through a constantly replenished layer of oil on top of a column of water; this approach limited the amount of particulate oil that entered the water column (Moles & Rice 1983). A continuous flow of diesel fuel WSF was generated by adding oil to a flow of water with a peristaltic pump (Croce & Stagg 1997). Contact of water and oil was enhanced by a receiving tank with serrated baffles, and a submersible pump distributed the WSF to a separator tank to remove particulate oil before distribution to exposure tanks for toxicity testing. Nordtug et al. (2011) pumped a mixture of oil and sea water through in-line nozzles to create a continuous dispersion of small oil droplets that promoted oil-water partitioning of hydrocarbons. The solution was filtered to remove oil droplets from the WSF and to prevent surfacing of oil droplets and adhesion to wetted surfaces. The system generated predictable and controlled oil-in-water dispersions and WSFs. These continuous-flow systems are less commonly used than static or semi-static systems, likely because they are technically more complex and more expensive. Nevertheless, they are more likely to generate data that are comparable among tests because there is less change in solution chemistry during testing.

The mixing instrument and the energy it transmits will influence the size distribution of oil droplets, their adhesion to equipment or surfaces, and the concentration of dissolved and particulate hydrocarbons in test solutions. Oils with different viscosities and composition, variations in oil loading, and variations in mixing energy will influence droplet size distribution and the ratio of dispersed to dissolved oil (Nordtug et al. 2011). A wrist-action shaker, a centrifugal pump, and a commercial blender were used to mix oil and water to assess the size distribution of oil droplets in mechanically-generated emulsions (Stokes and Harvey 1973). Based on droplet size and distribution, the authors selected the commercial blender as the standard instrument to prepare test solutions. To produce a WAF with less than 5 mg/L of oil droplets, filtration systems must be capable of removing droplets in the 2 to 3-micron range.

Publications that provided a diagram of their mixing system or instruments, especially if novel or complex, were helpful in interpreting and reproducing the results. The mixing instrument/system used to generate test solutions was often chosen based on its perceived mixing energy. Nevertheless, without methods to measure mixing energy, or quantitative relationships between mixing energy and droplet size, there are no tools to compare the effectiveness of different mixing instruments or systems. Gardiner et al. (2013) was one of the few papers that compared methods with different energy, but the difference was created by applying periodic mechanical rocking of a 20 L aspirator bottle. The rocking was compared to slightly energetic wave conditions (sea state 2-3), but no quantitative mixing energy was determined. The only way to assess the effectiveness of different mixing systems is to compare the characteristics of solutions that are generated, including size distribution of oil droplets, hydrocarbon composition and concentrations, and toxicity (see section 4.3.3). However, measurements of droplet size and hydrocarbon concentrations are expensive and not often reported with toxicity test data.

The physical characteristics of the test oil also play a critical role in the effectiveness of oil-water mixing. As outlined in section 4.1.1, highly viscous oils are difficult to disperse and produce large and unstable oil droplets, resulting in the slow partitioning of hydrocarbons from oil to water (Martin et al. 2014). Their inherent toxicity may be underestimated because the most toxic constituents may not partition into test solutions to toxic concentrations on the test time-scale. Because weathering of oil increases viscosity, comparisons of toxicity between unweathered and weathered oils may be biased by a reduced effectiveness of physical dispersion. This is particularly relevant to tests with dilbit, as the fraction remaining after weathering is similar in density and viscosity to a heavy fuel oil (ENG 2011 as quoted in GOC 2013), and weathering of dilbit stocks can occur during a program of toxicity testing.

Priority Research Needs

- Develop a routine method (or system) for determining mixing energy that can be followed in practice by any researcher – including sonication and high energy mixing.
- Conduct a literature survey to demonstrate the range of mixing energies associated with spills to open oceans, coastlines, lakes, and rivers of different current strengths and turbulence.

Therefore, standardized methods for preparing solutions of weathered dilbit could include application of chemical dispersants, increased mixing energy, or partition controlled delivery to overcome the high viscosity. For risk assessments, perspective is needed on the toxicity of dilbit under actual spill conditions, which could be obtained by repeating tests with lower mixing energies and by quantitative measures of mixing energies in real environments and in the lab.

4.2.3 Diluent water for preparation of stock and test solutions

The use of fresh versus sea water as a diluent will affect the behaviour of oils in solution, particularly the density and buoyancy of droplets. As sea water is denser (1.024 g/cm^3) than fresh water (0.998 g/cm^3 at 20ºC), oil added to sea water will be less likely to sink or become entrained in the water column than oil added to fresh water. For dilbit, the effect of oil weathering (48 h at 15ºC) was to increase its density to greater than that of fresh water (suggesting sinking) but less than that of sea water (suggesting floating) (GOC 2013). Unfortunately, there are few studies of the behaviour of oils in fresh water (Dew et al. 2015), and more research is needed.

Salinity also affects the accumulation and toxicity of hydrocarbons to aquatic organisms because PAHs are less soluble in salt water than fresh. Ramachandran et al. (2006) found that PAH uptake by juvenile mummichog and rainbow trout exposed to oil decreased as salinity increased from 0 (fresh water) to 15 (brackish) and 30 g/L (sea water). Thus, risks to fish from spilled oil should be greatest in fresh and brackish water and lowest at high salinity. Uptake of soluble PAH by fish from petroleum products may also be sensitive to changes in gill permeability and osmoregulation. Shukla et al. (2007) observed the role of salinity on uptake of PAH from crude oil by *Tilapia mossambica* at a salinity range of 0 to 33 g/L and found that concentrations of aromatic hydrocarbons in water of low salinity (i.e., coastal waters or estuaries) were higher than in areas of high salinity (i.e., open ocean) and tilapia had higher PAH uptake in lower salinity waters.

Salinity also has implications for dispersant effectiveness and toxicity. For example, Corexit 9500 effectiveness decreased as salinity increased (Ramachandran et al. 2006), but dispersant toxicity increased with salinity (Wilson 1977). Over a range of salinities, both Corexit 9500 and Corexit 9500-oil mixtures remained toxic to estuarine fish for at least four weeks at the lowest tested salinity (4 g/L), suggesting higher sensitivity of test species or reduced biodegradation of toxic components in low salinity environments (Kuhl et al. 2013).

CROSERF provides guidance for the type and quality of diluent water for toxicity tests (Aurand & Coelho 2005, Appendix 1). Under these protocols, tests should use filtered (0.45 µm) natural sea water diluted with deionized water as necessary to achieve a specific salinity. If natural sea water is not available, re-constituted sea water known to consistently and reliably support good survival and health of test organisms could be used. Studies comparing natural to artificial diluents are scarce, but Echols et al. (2016b) and Freeman (1953) recommend the use of synthetic diluent water with distilled water and inorganic salts because it is repeatable between laboratories. The specification of diluent water as sea water indicates that CROSERF protocols were designed for marine species and applications, particularly in response to large volume marine oil spills.

Freshwater tests may require modification for optimal results. For example, oil spills to fresh water often involve lakes or rivers that are turbid due to high levels of primary productivity (algal cells) or suspended inorganic particulates due to erosion. Few publications reported suspended solids in diluent water, primarily because most use laboratory supplies of filtered water. Spilled oil can adhere to fine clays, forming oil mineral aggregates (OMAs) and floating tar balls (Lee et al. 2015). While it is unlikely that OMAs will increase toxicity, they may cause toxicity to be underestimated if they decrease the bioavailability of the test oil. They may also increase the exposure duration by increasing the persistence of the oil in the exposure system. Studies of oiled sediments were excluded from the scope of this literature review, but there is a clear need for sediment toxicity tests to support freshwater ERAs. Humic substances also occur at high concentrations in some fresh waters, and their effect on the dispersion and fate of oil has not been evaluated.

For all 144 publications reviewed, the diluent water for oil/water stock solutions and the water used in test solutions were designated as either **Fresh water** or **Sea water** (Figure 6); waters were not differentiated by salinity within the sea water category. The salinities of brackish water range from 1-10 g/L (National Ground Water Association 2010) to 0-35 g/L (NOAA 2008). Therefore, publications that reported a salinity greater than 1 g/L for diluent water for test solutions were classified as sea water, including water that could be considered as brackish or estuarine. If both fresh and salt water were used, the paper was designated as **Both**. If the paper did not explicitly specify the water type, the water type was classified as Not Specified (**NS**). Diluent for test solutions and diluent for stock solutions were commonly from the same source and given the same designation. When diluent water was referred to in a paper as “embryo rearing medium”, the entry was designated as fresh water or sea water depending on the natural habitat of embryos of that study.

The dilution water used to prepare both oil/water stock solutions and test solutions was primarily sea water (Figure 6). The remaining publications reported freshwater diluent, including municipal and distilled water sources, but many failed to explicitly specify dilution water or to provide sufficient information that it could be identified. More papers specified dilution water for stock solutions than for test solutions, likely because the preparation of stock solutions is described first and the reader must assume that the same water type was used for test solutions. As with “embryo rearing medium”, the dilution water for test solutions could sometimes be determined by the preference of the test organism or some other designation in the methods description. The salinity of water used to prepare stock solutions was specified by

33 publications and 49 reported the salinity of the test solutions. Salinity of dilution waters ranged from 15 to 35 g/L, most commonly from 20 to 33 g/L, and was most frequently reported (21.2%) as 30 g/L. The salinity of toxicity test solutions ranged from 3 to 39 g/L, with the most common within the range of 20 to 35 g/L, and the most frequently reported (14.3%) being 30 g/L.

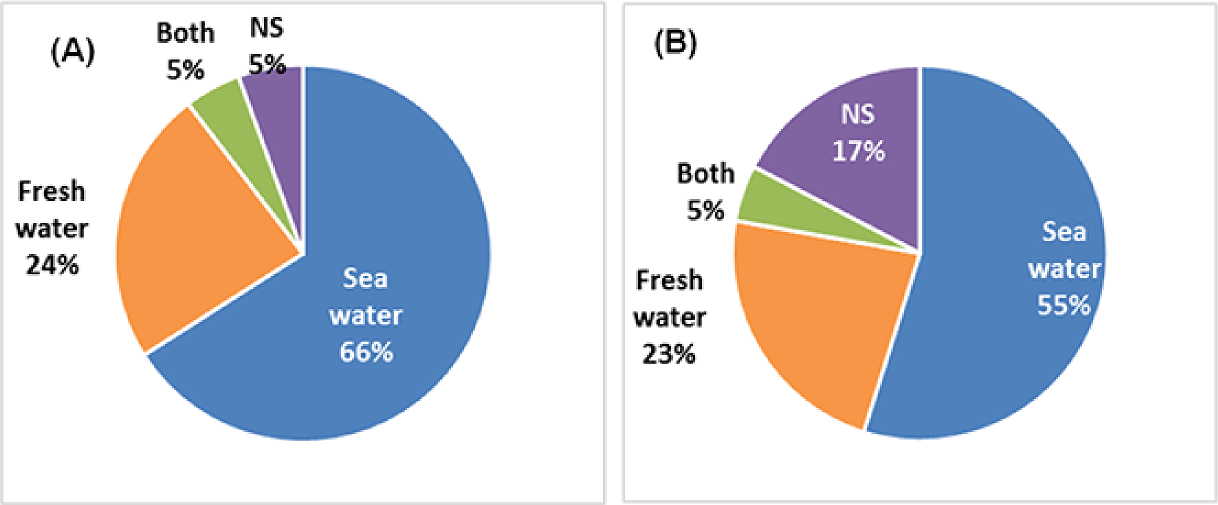

*Figure 6: The percentage of tests (N = 144) that used fresh water, sea water, both, or did not explicitly specify (**NS**), for diluent when preparing stock oil/water mixtures of WAF or CEWAF (A) and when preparing test solutions (B).*

Fluctuations in salinity may add an element of realism to a study. Approximately 5% of the publications reported both fresh water and sea water (**Both**) for preparing stock and test solutions, often to simulate estuarine or coastal environments. For example, Carls and Thedinga (2010) increased salinity twice daily for one-hour periods to simulate salinity changes during natural tidal cycles. Similarly, Heintz et al. (1999) used a combination of fresh and sea water to simulate intertidal environments. Changes in salinity may also be used to mimic seasonality, for instance as ice melt and evaporation will affect the salinity of a given aquatic environment.

Research Need

- Assess the effects on oil fate and behaviour of high concentrations of dissolved and particulate organic carbon, particularly humic acids, in fresh water.

4.2.4 Oil-to-water ratio (OWR)

The mixing volumes of oil and water are commonly reported as oil-to-water ratios (OWR). The composition and concentration of the test solution is a function of the amount and relative proportions of hydrocarbons in the oil and the unique partitioning rates of each component of oil into water. As a result, the variable loading method (variable OWR) recommended by CROSERF will generate test solutions with different concentrations and proportions of hydrocarbons. For example, Shiu et al. (1990) measured the solubility of 42 crude oils at different temperatures (5 to 22˚C) and salinities (double distilled water, 0 to 3% sodium chloride) with oils of varying weathering state and OWRs. The composition and concentration of WSF was dependent on the characteristics of the parent oil and the OWR. Crude oil WSFs prepared with an OWR of 1:5 contained higher concentrations of hydrocarbons which were dominated by BTEX (80%) than WSFs prepared with an OWR of 1:100 000, which contained only 15-50%

BTEX. The composition of WSFs prepared from more viscous bunker oils with a lower fraction of volatile components was more consistent over a range of OWRs. For other oils, there was a threshold OWR above which no additional changes in WSF composition were measured. The threshold for PBCO was < 1:1000 and for No. 2 fuel oils < 1:20. Shiu et al. (1990) concluded that test solutions prepared with different OWRs cannot be compared as the composition and concentration of the aqueous phase varies with the OWR. When mixing energy, salinity, and temperature were held constant, the concentrations of hydrocarbons in the WSF of Arabian crude and lubricating oil increased with the OWR, but to a different extent for each oil and not proportionally; similar results were reported by Maher (1982; 1986).

Several methods have also been reported for adding oil to water, including adding oil by weight to water that is already mixing with a vortex (Gardiner et al. 2013) or by volume from a gas-tight Hamilton syringe (Tjeerdema et al. 2007). No publications were found that compared the chemical composition or toxicity of test solutions generated by different methods of adding oil.

The OWRs were often reported in each publication as a ratio of oil to water. If the weight of oil to volume of water was reported, the ratio was calculated. Both CROSERF (Aurand & Coelho 2005) and a commonly-referenced modified method (Singer et al. 2000) report the addition of oil to water in ranges as they recommend variable loading to produce test solutions; CROSERF recommends 0.01 to 25 g/L, while Singer et al. (2000) used 0.01 to 10 g/L. When the weight of oil added was reported instead of volume, the OWR was calculated assuming an oil density of 1.0 g/mL unless otherwise stated.

The frequency with which each OWR was reported was graphed (Figure 7), but if an OWR was reported three or fewer times, it was classified as **Other**. If the amount of oil and/or water was not explicitly stated in the text, the OWR was recorded as Not Specified (**NS**). If oil was not added to water (e.g., oil was added to a substrate), the entry was classified as **NA** (not applicable) and not included in the analysis.

In the literature reviewed, OWRs ranged from 1:106 667 to 1:1 (Figure 7). Of 219 entries (some publications tested multiple ratios), 157 OWRs were reported, 44 of which were unique. Of the 157 reported OWRs, the most frequently reported was 1:9, (21.0%), a ratio originally reported by Anderson et al. (1974) that pre-dated CROSERF guidelines. Fifty eight entries were classified as **Other** (26.5%), indicating a diverse array of methods that confound comparisons of oil toxicity among studies.

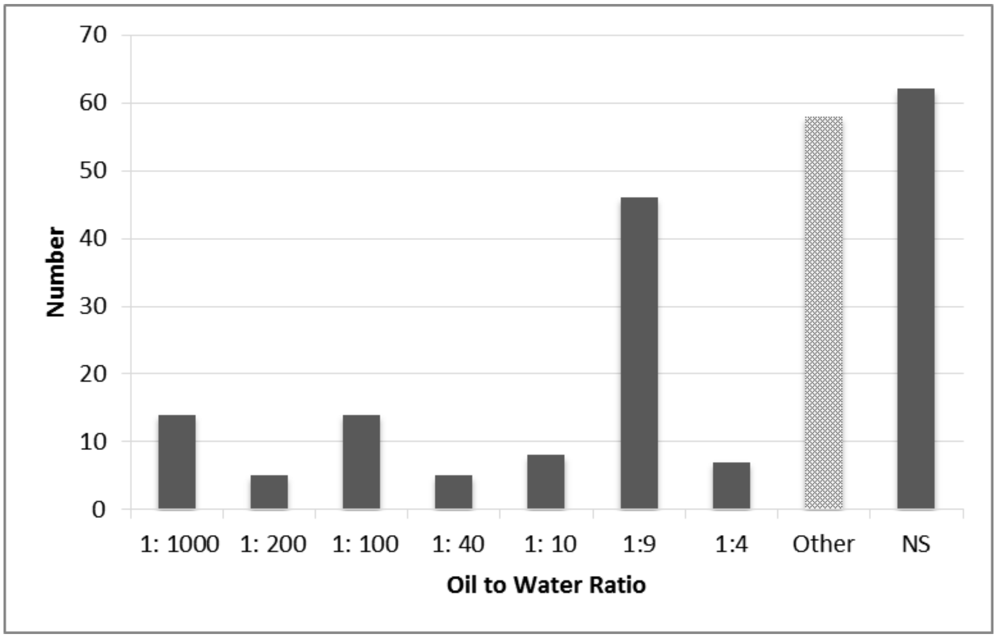

Figure 7: The number of oil:water ratios (OWR) reported in oil toxicity tests (N=219). Some studies used multiple ratios in a single study. Ratios reported 3 or fewer times were classified as ***Other****. When the volume of oil and/or water was not reported, the publication was classified as* ***NS*** *(not specified). 23 publications were not included because they did not involve adding oil to water.*

Given the influence of OWR on the chemical composition of test solutions, it should be included in all test reports.

4.2.5 Mixing vessel size and shape

Mixing vessel size should be determined by the approach to preparing test solutions, by the volume of solution required for the toxicity test, and by the mechanics of mixing. For tests with fish embryos (often <10 mg each), the volume of test solution needed per day in each treatment may be less than 100 mL while tests with 10 juveniles weighing 10 g each would require 100 L per day (see section 4.4.3). Where a unique solution is created for each concentration of a test series (CROSERF approach), a mixing vessel equivalent to the size of the test chamber is required. Alternatively, if solutions are prepared by diluting a single stock of WAF or CEWAF, a much smaller vessel is needed to create a concentrated stock that may be more practical to create and dispense. CROSERF encouraged a range of 1 to 20 L to minimize the effects of logistics and geometry on the composition of test solutions. A modified CROSERF method commonly referenced by other publications (e.g., Singer et al. 2000), did not specify a mixing vessel size. Anderson et al. (1974) used a mixing vessel volume of approximately 19 L.

Mixing vessel geometry may affect the partitioning of hydrocarbons from oil into water. The surface area-to-volume ratio of mixing vessels determines the thickness of oil films during mixing and the surface area for oil-water partitioning of hydrocarbons. Therefore, partitioning kinetics change with the thickness of the oil layer and not solely with the OWR (Tsvetnenko & Evans 2002; Ramachandran et al. 2004). There are few systematic assessments of the effects of oil layer thickness on hydrocarbon concentrations in water.

It is recommended that the mixing vessel size strike a balance among what is practical in the lab, an oil thickness that allows for maximum partitioning of hydrocarbons between oil and water, and the volume of test solution needed for each toxicity test. The size and biomass loading of test organisms is an important factor dictating the volume of test solutions (Section 4.4.3) and must be chosen with care when planning a test.

Each of the publications that reported a mixing vessel volume was given a designation for mixing vessel size as a volume in liters (N=201; Figure 8). Volumes were placed into ranges rather than individual values to limit the number of categories. Most publications used 1 L to < 100 L mixing vessels, but 38.3% did not specify (**NS**) the volume. If oil and water were not mixed in a vessel (i.e., if oil was added to a substrate as opposed to mixed with water), the method was designated as Not Applicable (**NA**). The larger volumes used for mixing oil and water reflect the CROSERF protocol (Aurand & Coelho 2005) in which each oil-water concentration in a toxicity test is mixed independently. The smaller volumes usually reflect the preparation of a single stock solution with subsequent dilutions.

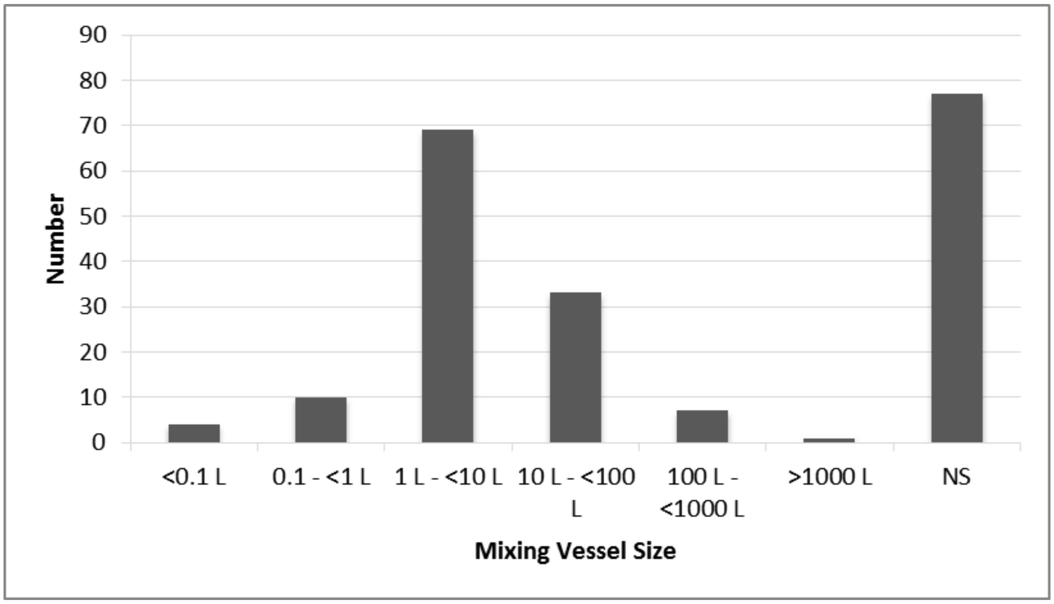

Figure 8: The number of vessel sizes used to mix oil and water for preparing toxicity test solutions, ***NS*** *where the mixing vessel size was not specified (N=201).*

4.2.6 Headspace in mixing vessel

The headspace is the portion of the mixing vessel volume above the liquid aqueous phase. The headspace is relevant when mixing vessels are closed during the preparation of WAF or CEWAF; vessels covered with tinfoil were considered as open vessels. Headspace and surface area-to-volume ratios were among the most significant characteristics of a mixing vessel (Singer et al. 2000). A closed headspace will influence the composition of the oil-water test solution because it traps volatile compounds that may be acutely lethal and limits their loss from solution during oil-water mixing (Redman & Parkerton 2015). Thus, the influence of a closed headspace on the retention of volatiles is relevant to ensuring that tests of acute lethality of fresh oil are unaffected by the loss of volatiles during solution preparation (NRC 2005). Singer et al. (2000) noted better reproducibility of acute toxicity test results in test solutions when the headspace was 20% or less, likely due to differences among labs in test conditions that affect volatility (Aurand & Coelho 2005) (e.g., temperature). For chronic toxicity, fish are most likely exposed to weathered oil, so the loss of volatiles during mixing would provide a more relevant test solution. Loss of volatiles during solution preparation is a particular problem for unweathered dilbit products (Sections 2.1.1; 4.2.11), and should be minimized for more accurate tests of acute lethality, and to avoid changes in viscosity and dispersibility of dilbit during prolonged stirring. On this basis, the CROSERF guideline for acute lethality tests should be followed.

Approximately 10% of 144 publications reported the amount of headspace for mixing vessels. Because so few publications reported this value, it was excluded as a parameter for descriptive statistics in this review. Headspace volumes that were reported included none (Redman & Parkerton 2014; Hook et al. 2010), below 20% (Echols et al. 2016a; Bidwell et al. 2003), and 20 to 25% (Adams et al. 2014a; Adams et al. 2014b); CROSERF guidelines recommend a headspace of 20 to 25% (Aurand & Coelho 2005; Barron & Ka'aihue 2003). Some papers reported headspace as a ratio of water to air; a headspace ratio of 4:1 was reported in two publications (Faksness et al. 2008; Melbye et al. 2009). The lack of papers that detailed the headspace of the mixing vessel suggests that most studies did not consider it important, but clearly this is a characteristic of solution preparation that must be reported for both acute and chronic toxicity tests.

For unweathered dilbit it is recommended that the mixing vessel be closed with minimal headspace for retention of volatiles during mixing. Mixing of weathered products can be open or closed.

4.2.7 Chemical Dispersion: Types of dispersant

Spill treating agents are available to remove surface films of oil from water (dispersants), clean oil from shoreline substrates, cause oil to gel so that it may be more easily removed from a spill site, and promote the ignition and combustion of surface oil to remove oil slicks by burning. The CROSERF protocols were developed specifically to test the toxicity of chemical dispersants and dispersant-oil mixtures, and virtually all of the literature on the toxicity of spill treating agents reviewed for this report described tests with dispersants.

Chemical dispersants are mixtures of surfactants, additives, and solvents that reduce the interfacial tension between oil and water, facilitating the formation of oil droplets and stabilizing droplets to limit re-coalescing as surface films. The overall effect is to increase the rate of oil dilution in the water column (Clark et al. 2005). For example, chemical dispersions of heavy fuel oil increased the concentration of TPH in CEWAF by 25-fold compared to WAF (Martin et al. 2014). The amount of oil dispersion depends on the characteristics of the oil, the conditions of the test, and the type of dispersant, which also affects the size distribution of oil droplets (Clark et al. 2005). These criteria have been used to screen and develop an acceptable list of dispersants for use in Canada (Doe & Wells 1978).

There are a wide variety of chemical dispersants available, and the dispersants captured in this report typically represent those easily attainable and approved for use within the geographic region of the study. The dispersants most frequently used were Corexit 9500 and Corexit 9527, reported in more than 50% of studies of chemical dispersion. This frequency reflects the dispersants recommended for use by CROSERF (Aurand & Coelho 2005; Barron & Ka'aihue 2003) and that most studies assessed were conducted in North America, where Corexit dispersants are approved for use. The name of the chemical dispersant used was not specified in 4% of publications, and all others were used less than four times, including:

- Berol TL 188
- Berol TL 198
- BP1100
- BP1100X
- Corexit 19-L-50
- Corexit 7664
- Corexit 8666
- Corexit 9517
- Crystal Simple Green
- Dasic Slickgone NS
- Dispersit SPC 1000
- Drew Chemical OSE 71
- Drew Chemical OSE 72
- Finasol OSR-2
- Finasol OSR-52
- Finasol SC
- HiClean II
- Hytron #3

- Inipol IP 90
- Innospech
- JD-2000
- Linco-6
- Max Clean 2
- Nafleet Oil Spill Dispersant
- NEOS AB3000
- Nokomis 3-AA
- Nokomis 3-F4
- Oilsperse 43
- Polyclens Industrial TS7
- Saf-Ron Gold
- Sea Brat 4
- Sea Green 80
- Shell dispersant LTX
- Shuangxiang No.1
- Sugee #2
- Symperonic OCD 20
- Total Fluides
- Weipu MD -55
- ZI-400

The toxicity of dispersants has been compared under varying conditions, including temperature, salinity, and exposure duration. Corexit 9527 and 9500 are considered to be low to moderately toxic; 24-96 h LC50s and EC50s range from 1.6 to > 1000 mg/L and 0.7 to > 400 mg/L, respectively (George-Ares & Clark 2000). Laboratory methods to assess the toxicity of dispersant have been criticized for exposure durations and static renewal regimes that are not consistent with dispersant applications during actual spills (George-Ares et al. 1999; George-Ares & Clark 2000). However, with the continuous and prolonged use of Corexit 9500 to control the Deepwater Horizon oil spill, experiments with prolonged exposures to dispersant and dispersed oil are actually quite relevant. Dispersants are also important experimentally as tools to remove the confounding effects of viscosity in comparisons of toxicity among oils of widely differing properties (e.g., Adams et al. 2014a).

The difference in toxicity between Corexit 9500 and 9527 can be attributed to variations in solvent composition. Corexit 9500 was specifically created for greater applicability under a variety of conditions and oil types (George-Ares et al. 1999; George-Ares & Clark 2000) and is more effective over a broader range of salinities than Corexit 9527 (Blondina et al. 1999). The peak dispersion efficiency for a suite of oils with dispersants, including Corexit 9526, Enersperse 700 and Citrikleen, was at 40 g/L salinity with efficiency declining as salinity decreased to 0 g/L or increased to 80 g/L (sea water is 32 g/L) (Fingas et al. 1991). Effectiveness decreased at dispersant-to-oil ratios greater than 1:40 to 1:60. In general, higher mixing energies increased dispersion effectiveness, oil weathering decreased dispersant effectiveness, and at each weathering state, effectiveness increased with salinity (Chandrasekar et al. 2006; Clark et al. 2005).

Consistent with the trends in Fingas et al. (1991) and Chandrasekar et al. (2006), Blondina et al. (1999) found difficulty in dispersing heavy oils and observed that the oil remaining after chemical dispersion consisted of stringy mats. Dispersion of a range of viscous heavy fuel oils was reduced for oils with a higher n-paraffin content, likely due to the cohesiveness of the paraffin molecules (Canevari et al. 2001). More viscous heavy oils also resist dispersion due to slower rates of dispersant penetration (Lessard & DeMarco 2000; Kaku et al. 2006). Because dispersant formulations were manufactured to have maximal dispersion in salt water (Chandrasekar et al. 2006), their effectiveness is likely lower in fresh water (Lee et al. 2015). No studies reported the effectiveness of chemical dispersion of unweathered or weathered dilbit. Although there is a wide range of dispersant toxicities and effectiveness, the choice of dispersant depends more on availability and relevance to permitted uses than to optimizing toxicity test methods.

Chemical dispersion for standardized tests may be beneficial to maximize partitioning from oil into water without changing the toxicity of the oil (as was observed with viscous heavy oils). Dispersant effectiveness will vary with the weathering state of dilbit and should be assessed prior to toxicity testing. The only published study with chemically-dispersed dilbit did not assess the effectiveness of dispersion of CLB with Corexit 9500 (Madison et al. 2015).

4.2.8 Chemical Dispersion: Dispersant-to-oil ratio

The dispersant to oil ratio (DOR) is a ratio of the volume of dispersant applied and the volume of oil in the mixing vessel. It is important to choose an appropriate DOR as over-dosing can lead to surplus dispersant in the water column that can be acutely toxic, and under-dosing can reduce dispersant effectiveness. Heavy oils or oils spilled in calm water require a greater DOR compared to light oils or spills to turbulent water where mixing energy is high (Chapman et al. 2007). The DOR depends on a broad range of factors, including natural dispersibility and weathering of the oil, settling time, mixing energy and even the corresponding oil-to-water ratio (Clayton et al. 1992).

There is still debate about the conditions in which chemical dispersants cause toxicity in dispersed oil toxicity tests. Couillard et al. (2005) noted that dispersants increased CYP1A activity in fish and caused greater CYP1A-mediated toxicity. Exposures of Japanese medaka to dilbit CEWAF caused greater molecular responses than dilbit WAF (Madison et al. 2015). Some of the difference could be attributed to higher concentrations of TPH in CEWAF compared to equivalent dilutions of WAF, but some molecular responses were unique to the Corexit 9500 dispersant, as indicated by dispersant controls. Therefore, dispersants in test solutions may contribute to the toxicity of dispersed oil mixtures (see section 4.2.9).

The effectiveness of chemical dispersion is measured by the extent to which oil slicks are broken into smaller droplets and entrained in the aqueous phase. Modification of the DOR changes the interfacial tension of the oil, and the highest dispersion occurs with the maximum reduction in interfacial tension (Khelifa et al. 2007). The DOR commonly used for oil toxicity tests may not be the DOR which produces the smallest droplets or the optimum/maximum dispersion under the test conditions, and a higher DOR does not always cause a greater reduction of interfacial tension. For example, dispersions of ANSC oil with Corexit 9500 at a DOR of 1:20 produced smaller oil droplets after 90 minutes of high energy shaking than dispersions with a DOR of 1:10. In contrast, a DOR of 1:10 was more effective at dispersing IFOs than DORs of 1:25 and 1:50 (Clark et al. 2005). Methods for adding dispersant to oil in the mixing vessel were rarely reported, but have an important role in dispersant effectiveness. Methods included adding the dispersant “in accordance with the instruction from manufacturer” (Adeyemo et al. 2015), adding during mixing (Agamy 2013), pre-mixing oil and dispersant before application (Oslvik et al. 2012), adding dispersant "drop by drop" to the oil (Yu et al. 2015); adding dispersant to the centre of the vortex of a stirred solution of oil on water (Adams, unreported data) or distributing the dispersant over the entire surface of the oil layer (B. Madison, Queen’s University, personal communication). Comparisons of the toxicity of chemically-dispersed oils with different physical characteristics may be biased if there are differences in dispersant effectiveness, so *it is critically important that DORs and dispersant application techniques be reported with toxicity test methods*.

Of the 226 mixing methods reported, 66 used chemical dispersion to generate test solutions (see section 4.2.7). Each of the 66 methods was classified by DOR; five publications tested multiple DORs, giving a total of 74 entries with 68 reported DORs. When publications reported only the amounts used, the DOR was calculated, assuming a density of 1.0 where dispersant weights were reported. Six publications did not report the DOR or amounts of dispersant used and were classified as **NS**. DORs ranged from 1:500,000 to 1:1, and of the 74 entries, 37.8% used a DOR of 1:10, 17.6% used a ratio of 1:20, and 8.1% used a ratio of 1:30 (Figure 9). A separate category (**Other**) was used to collect data for any ratio reported that had less than or equal to three entries and included DORs of 1:500 000, 1:2000, 1:1200, 1:1000, 1:100, 1:50, 1:40, 1:25, 1:9, 1:5, 1:3, and 1:1.

There are no publications to date on the optimal DOR and mixing energy needed to generate chemically-dispersed solutions of dilbit for toxicity testing. If dispersants are used to minimize the effects of viscosity on the dispersion of dilbit, the DOR and mixing energy should be optimized to produce the maximum dispersant effectiveness. If dispersants are applied to mimic dispersant use at actual spills, the mixing energy and DOR should be consistent with the observed dispersant use at the spill site. The maximum ratio for spill sites is 1:20 based on spill response compliance with application rates published on dispersant product labels (ASTM STP 1282; proposed framework 4.6). For tests with dilbit, research is needed on the trade-off between dispersant effectiveness and toxicity, and on the size distribution of droplets produced at different stages of weathering. Addition of dispersant to a greater surface area of the floating oil may enhance the dispersion of dilbit and more viscous oils.

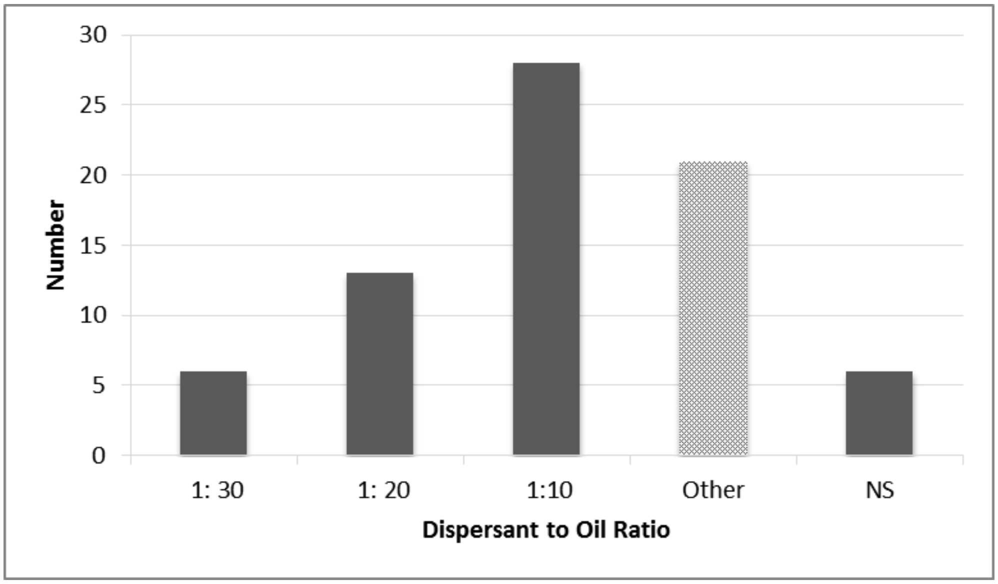

*Figure 9: The number of the most commonly reported dispersant-to-oil ratios (DOR). If a given DOR was used three or fewer times among publications, it was classified as **Other**. If the DOR was not given or could not be calculated using information provided in the publication, it was classified as not specified (**NS**).*

Research Needs

- Identify the optimal dispersant-oil ratios and methods for adding dispersant to oil to minimize ‘free’ dispersant in test solutions, and to maximize the amount of dissolved and particulate oil in solution. For risk assessments, tests of different ratios may be needed to match site-specific conditions or applicable regulations.
- Assess different controls for tests of chemically-dispersed oil to identify whether or not dispersant affects the outcome of toxicity tests, and whether there is oil-dispersant interactive toxicity.
- Assess the effectiveness of chemical dispersion of weathered and unweathered dilbit according to the size distributions of droplets produced, the chemical composition of test solutions, and measured toxicity.

4.2.9 Chemical Dispersion: Dispersant Controls

The potential for synergistic toxicity of dispersant and dispersed oil has been the topic of much debate and highlights the importance of controls in oil toxicity testing. In addition to positive controls to detect effects and negative controls to assess the health of unexposed test organisms, controls for any solvents or methods used should be included. This is particularly important for toxicity tests of chemically-dispersed oil to determine whether there is a contribution from free dispersant in the test solution. Dispersant controls may include chemical dispersions of non-toxic mineral oil (e.g., Nujol) to distinguish between the toxicity of the dispersant and oil (Adams et al. 2014b). In the studies reviewed that involved chemical dispersion, half included a control toxicity test with dispersant alone, of which a small percentage reported a dispersant-mineral oil control (4.8%).

Alternatively, the potential toxicity of the dispersant in the mixture can be estimated by measuring unbound or ‘free’ dispersant in CEWAF stock solutions to compare with an exposure-response dataset of the dispersant alone. This method assumes additive toxicity of the dispersant and oil components in the CEWAF treatments. Analytical methods are available to quantify the amount of dispersant in toxicity test solutions. For example, components of Corexit 9500 and dioctyl sulfosuccinate have been measured by high-resolution mass spectrometry (HRMS) and Charged Aerosol Detection respectively (Choyke & Ferguson 2015). Aurand and Coelho (2005) outlined the CROSERF protocol to measure dispersant with UV-VIS spectroscopy. However, it is unknown whether these methods can distinguish the ‘free dispersant’ from dispersant associated with oil droplets in CEWAF samples.

Future experiments with chemical dispersion should use one of these types of controls to identify any dispersant-related effects or synergism with oil toxicity.

Research Needs

- Assess the chronic and sublethal effects on fish of chemical dispersants, both alone and in combination with oil.
- Develop methods to characterize ‘free’ dispersant in test solutions and dispersant associated with oil particles.

4.2.10 Mixing conditions

A variety of mixing conditions can affect toxicity test results and the dispersion of oil including water quality (pH, water hardness, salinity), temperature (Girling 1989) and light conditions. When mixing oil and water, the water pH can promote the formation of stable and viscous emulsions, most notably at neutral and basic pH (Daaou & Bendedouch 2012). Water hardness, specifically high concentrations of minerals, can also affect mixing of water and oil, but highly saline waters favour water and oil mixing compared to increasingly low salinities (Austad et al. 2011). Overall, water quality (pH, water hardness, organic content, salinity) was reported infrequently among the studies reviewed, and summary statistics were not prepared for this review. It is not possible to specify ‘standard’ water quality parameter values or ranges because the appropriateness of parameters depends on multiple factors, including the scenario of interest, the species to be tested, source water available, and possibly specific details of the toxicity test method.

A wide range of mixing temperatures was reported across the studies surveyed. Temperature affects the solubility of oil in water, mixing effectiveness, and the rate of loss of low molecular weight volatile compounds (Shiu et al. 1990; Maher 1982). Tests with Arctic species often match the mixing temperature to the environment of the target species (e.g., Barron & Ka’aihue 2003; Gardiner et al. 2013). Lower mixing temperatures not only change mixing behaviour but also prolong equilibrium times due to slower rates of oil-water partitioning. Alternatively, mixing performed at warmer temperatures causes a greater loss of volatiles and associated changes in viscosity (Hemmingsen et al. 2005). Hence, there is a trade-off between realism (mixing at environmentally-specific temperatures) and standardization (mixing at room temperature). Mixing at room temperature will provide more reproducible stock solutions which can be diluted into test solutions at other temperatures according to the test species, but the effects of the temperature change during dilution on oil-water partitioning are not well characterized. Stock solutions mixed at other temperatures matching test solutions will avoid temperature changes and more closely mimic conditions at a particular spill site, but results will be more difficult to compare with studies at standard room temperature or other temperatures. *Studies with mixing of oil and water stock solutions at hotter or colder temperatures should include a control mixed at room temperature to indicate whether mixing temperature affects toxicity*.

Of 144 papers reviewed and 150 specified temperature entries, 118 entries reported a mixing temperature of 20-25ºC. If a paper did not report a mixing temperature, it was assumed to be room temperature (21ºC). As a consequence, the number of studies at this mixing temperature may be over-stated. Eight studies used 25-30ºC as a mixing temperature. Of 26 papers that reported mixing temperatures below 20ºC, temperatures ranged from 0ºC (N = 1; Davison et al. 1992) to 18ºC (N=5; Camus et al. 1998; Ramachandran et al. 2004; 2006; Shukla et al. 2007; McIntosh et al. 2010). Ten of these 26 publications were oiled substrate systems where the temperature was set to the optimal temperature of the test species.

Irrespective of the temperature regime used in a study, it is recommended that method details, including temperatures of all solutions and any schemes for changing temperature from one step to the next, be carefully reported.

Photodegradation and photooxidation studies have shown that light can affect oil composition and toxicity, but usually during the toxicity test after mixing has occurred (Section 4.4.2). However, illumination during the preparation of WSF increased the naphthalene fraction (Østgaard & Jensen 1983) and the preparation of WSF under sunlight increased the dissolution of oil and generated peroxides (Ziolli & Jadim 2001). According to the CROSERF guidance document, mixing should be carried out in darkness but only 26 of the 144 papers reviewed reported light conditions during mixing, of which 18 were mixed in darkness, 2 used LED lights, 4 used a photoperiod of 6 hours of light followed by 16 hours of darkness, and 2 studies conducted mixing in 10 hours of light followed by 14 hours of darkness. *It is recommended that solutions be mixed in darkness or under a low intensity light that does not include UV wavelengths*.

Research Needs

- Determine the sensitivity of results to variations in key water quality parameters.
- Examine the effect of mixing stock solutions and diluting into test solutions at various temperatures, including the impacts on chemical composition and toxicity.

4.2.11 Mixing duration

The mixing duration is the length of time (hours) that oil and water, or oil, water and dispersant, are mixed. The mixing duration determines how much oil partitions into the water phase. There are various factors that govern mixing time, which change the nature of the oil and stock solution involved. The time to equilibrium varies with the oil and the conditions of the test, and as a result a functional equilibrium may not be reached in a standard mixing time and toxicity may be underestimated. Faksness et al. (2008) found that the time to equilibrium is longer for WAFs of wax-rich oils and WAFs prepared at lower mixing temperatures (see section 4.2.10), likely due to increased oil viscosity and decreased solubility. This means more prolonged mixing is needed to transport a given amount of oil into the aqueous phase due to lower rates of oil-water partitioning. The time to equilibrium for lubricating oils was 24 hours, but depended on the oil and OWR (Singer et al. 2001a). For the 18 crude oils tested by Singer et al. (2001a) the WAF was similar for 24 to 48 hours for the given OWR, but they recommended that mixing durations should not exceed 24 hours to avoid bacterial action.

A value for mixing duration was given for every oil-water solution generated in each of the 144 papers (Figure 10); some papers used multiple oil-water mixtures or multiple mixing times for a single solution, resulting in 181 total entries. Oil and water mixing times varied widely, from less than 1.0 to >72 hours in the publications reviewed. The most frequently reported duration was 18 hours (21.4%), followed by 24 hours (17.6%) and 20 hours (12.1%). If the duration of mixing was not explicitly stated, it was classified as not specified (**NS**).

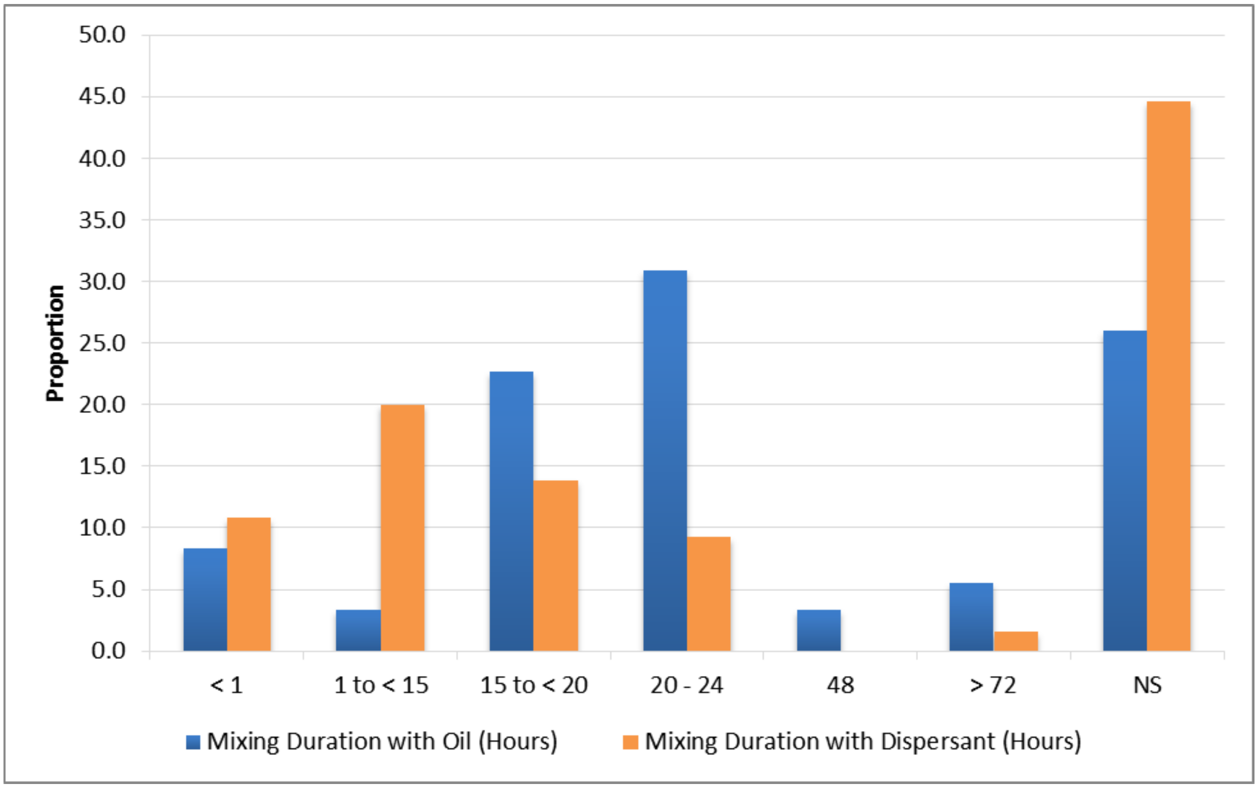

*Figure 10: The proportion of mixing durations (hours) of oil and water used to generate test solutions and mixing duration of oil and dispersant. If the publication text did not explicitly state the duration of mixing it was classified as not specified (**NS**). Mixing durations specified as a range were omitted from this graphic (N=1).*

There was a narrower range of mixing durations for oil, water, and dispersant solutions, from <1 to 24 hours. The most frequently reported duration was >1 hour, as reported by 27.7% of the papers, although 44.6% of the applicable papers did not state mixing duration. In comparison to oil and water systems, there is little literature regarding oil, water and dispersant mixing duration other than the CROSERF guidelines. The CROSERF guidelines suggest 18 hours of mixing, but timing should be optimized for each particular oil (Singer et al. 2001a). Another factor to consider is the daily cycle of preparing and renewing test solutions over periods of days (acute toxicity) to weeks (chronic toxicity). Mixing protocols that can be completed within 24 hours reduce the resources needed to run a test.

Dilbit undergoes rapid weathering during the preparation of test solutions due to prolonged exposure to air during stirring. The extent of weathering will depend on the nature and volume of diluent mixed with bitumen, and the temperature and length of mixing time needed to prepare stock solutions of WAF or CEWAF. The rapid weathering during handling and preparation of test solutions may change the viscosity and dispersibility of dilbit, the amount of hydrocarbons in solution, and the resultant toxicity. There are obvious interactions among the extent of weathering, the time to equilibrium for oil-water partitioning, the concentration of toxic constituents in water, and measured toxicity. *If the intent of a study is to measure the toxicity of fresh dilbit, stirring durations should be brief, even if this causes lower concentrations of oil in water. Reports on dilbit toxicity should describe how the properties of dilbit changed over the course of the sample preparation.*

Research Needs

- To develop optimum procedures for dilbit sample preparation.
- To assess the duration of mixing needed for dilbits with different formulations and weathering states to reach functional equilibria between water and oil phases, particularly across a range of temperatures.

4.2.12 Settling duration

The settling duration is the length of time (hours) that an oil-water (and/or dispersant) solution is allowed to stand un-stirred after mixing. As a sample settles, droplets of free oil may float to the surface, which reduces the proportion of free phase oil in test solutions, and reduces the error in chemical analyses of dissolved oil concentrations caused by sample contamination with droplets.

The total number of reported settling durations (N = 185; Figure 11) did not match the total numbers of methods (N = 223) because there are methods for which settling duration was not applicable. Settling durations ranged from 0 to 72 hours. The most frequently reported were one hour and six hours, reported in 26.4% and 11.9% of entries respectively. Settling duration was not specified (**NS**) in 29.7% of entries, and 11 settling durations (6%) involved a range of hours and are captured in the **Other** category. The range most frequently reported was three to six hours, consistent with CROSERF guidelines.

After settling, the bottom portion of test solutions was often decanted to separate it from the residual surface slick before toxicity testing. The CROSERF guidelines suggest that solutions be settled for 3 to 6 hours after stirring and then used immediately. Most concentration changes occurred within the first three hours of settling, and after six hours, there was often a decrease in variation among the mixtures in a gradient (Lunel 1993). As with mixing, it was suggested that settling duration not exceed 24 hours to avoid bacterial action (Singer et al. 2001a).

For dilbit products the optimal settling time will be highly dependent on the generation of oil droplets in water and their size distribution. Given the resistance to dispersion of weathered dilbit products under low energy mixing conditions, less settling time would likely be required compared to conventional crude oils; one hour was used by Madison et al. (2015) while Philibert et al. (2016) followed CROSERF (four hours).

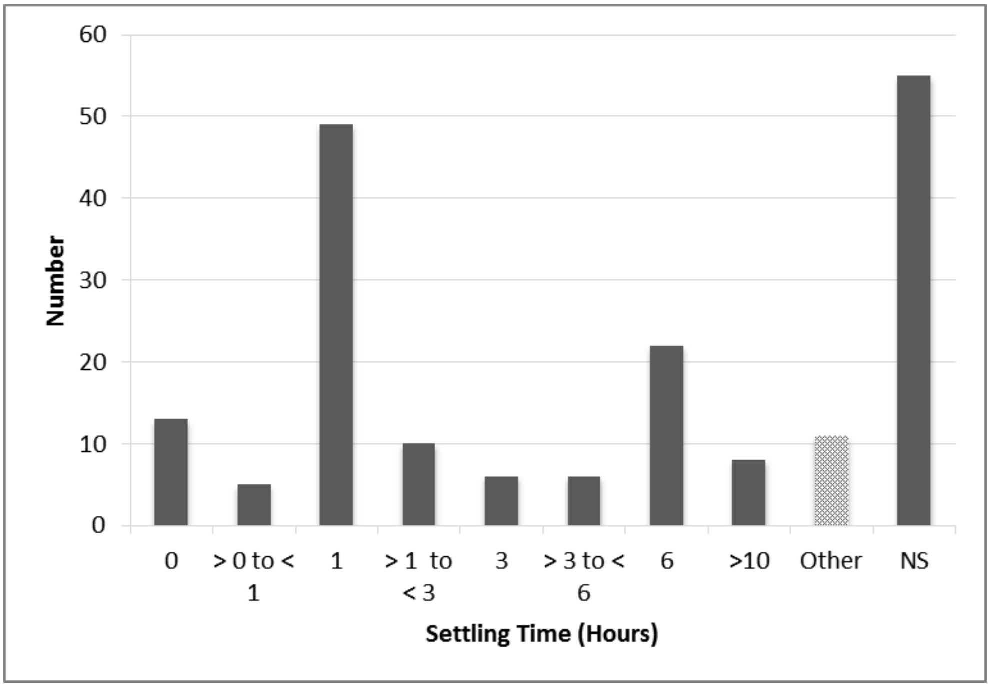

*Figure 11: The number of settling durations (hours) of oil and water used to generate test solutions. If the duration was not stated explicitly, it was classified as not specified (**NS**).*

4.3 METHODS TO CHARACTERIZE OIL AND TEST SOLUTIONS

Measured concentrations of oil in test solutions demonstrate whether test organisms were exposed to particular toxicants and to what concentrations (Harris et al. 2014). For oil toxicity tests, it is essential to measure the actual composition and concentrations of hydrocarbons in the parent oil and in test solutions. Because hydrocarbons are hydrophobic, it is also essential to measure temporal changes in concentration, over a 24-hour period for static or renewal protocols, and over the entire duration of the test. For studies comparing WAF and CEWAF, measured concentrations are also needed to support discussions regarding the use of dispersants (Clark et al. 2001).

One of the main goals of the CROSERF process was to improve the reproducibility of WAF and CEWAF solutions. The published CROSERF guidance included a list of chemical components that, ideally, would be measured in each WAF or CEWAF stock and diluted test solution (Singer et al. 2000; Aurand & Coelho 2005) (Appendix 1). Measurement of the complete CROSERF list of chemical components is normally divided into three analyses (Singer et al. 2000):

1. Total petroleum hydrocarbons (TPH) using gas chromatography with flame-ionization detection (GC-FID). This method typically detects hydrocarbons based on the number of carbon atoms, and covers the range C_{10} – C_{36}, though results extending below C_{10} and finishing below or above C_{36} are fairly common. It is possible to reconstruct the TPH parameter using individual hydrocarbon analysis results from GC with mass spectrometric detection (GC-MS), but this approach should be validated alongside GC-FID for a given oil before these TPH results are assumed to be equivalent.
2. Volatiles (often referred to as volatile organic carbons or VOC) by GC-FID or GC with mass spectrometric detection (GC-MS), where sample handling is specifically designed to reliably detect volatiles in the range C_6 – C_9. Currently, GC-MS is preferred over GC-FID. To ensure that all volatiles are captured in the analysis, Aurand and Coelho (2005) recommend

including cyclopentane, a C_5 compound with similar volatility to C_6 – C_9 hydrocarbons (Appendix 2, Table 2A). Volatiles usually include the monoaromatic BTEX (benzene, toluene, ethylbenzene, o,m,p-xylenes) compounds which frequently dominate the volatiles group. Some papers report only BTEX analysis to represent the whole VOC group.

3. Polycyclic aromatic hydrocarbons (PAH) by GC-MS, which is sometimes a subset of a "semi-volatiles" group. The sum of concentrations in this group is often called "total PAH" (TPAH). Aurand and Coelho (2005) list 40 aromatic compounds (or compound groups), which can be described as starting with the US-EPA list of 16 priority PAHs (all unsubstituted), adding biphenyl, benzo(e)pyrene and perylene, then adding dibenzothiophene (a heterocycle, but included for simplicity under PAH in most reports), and the C_1-C_4 alkylated homologs of naphthalene, fluorene, phenanthrene, pyrene, chrysene and dibenzothiophene. Many reports feature a smaller or larger group than this, as discussed below.

In general, the water-soluble fraction of oil includes the monoaromatic fraction (benzene, toluene, ethylbenzene, and xylene (BTEX)), other VOCs, and 2-ring PAH (naphthalenes). This fraction can be acutely lethal but is the least chronically toxic. It has been suggested in some studies of heavy and highly weathered oils that analysis of VOCs can be omitted as these will not be present in the source oil (Barron & Ka'aihue 2003). The intermediate-sized PAH including the 3- to 4-ringed alkyl PAHs are much less water soluble, but have been identified by EDCF as the main components of oil causing chronic embryo toxicity due to their bioavailability and environmental persistence (Carls et al. 1999; Hodson et al. 2007; Adams et al. 2014a). The correlation of toxicity with the alkyl PAH in oil and test solutions has been a driving force for the development of analytical methods to quantify these components in the oil, oil and water mixtures, and test solutions.

For dilbit oils, weathering during handling, solution preparation, and throughout the toxicity test may be more significant than for crude oils, mainly because of the volatile components of the diluent. This stresses the importance of chemical analysis to track changes in the physical and chemical characteristics of these products during the experiment. For dilbit products, monitoring the concentration, distribution and ratio of VOC to PAH in water samples may allow the tracking of dilbit chemistry with handling and testing over time.

4.3.1 Nominal versus measured concentrations

Most papers reviewed reported measured concentrations of hydrocarbons in test solutions (Figure 12). Only a small proportion (16%) did not report any analysis, and either gave no statement regarding concentration, or just reported dilution factors between a WAF or CEWAF stock and various test solutions. In a few cases where a WAF or CEWAF procedure was not used to disperse the oil, a concentration was stated based on the total mass of oil mixed with water, assuming complete dispersion. Concentrations based solely on dilution factors or assumed complete dispersion are referred to as "loading rates" (CONCAWE 1993), "nominal" loadings or "nominal" concentrations (Fucik et al. 1995). Rufli et al. (1998) recommended that nominal loading rates be described for comparisons of toxicity based on the amount of oil that gets into water. The main flaw of nominal loadings is the assumption that the protocol used would deliver the same concentration of oil into the WAF or CEWAF stock solution as was delivered in any other independent study where concentrations were measured (or that will be compared). As discussed above, this assumption is highly problematic even when the same oil is used, owing to the many factors that influence oil dispersion into WAF or CEWAF, including weathering factors that cannot be rigorously reproduced. These problems are magnified when comparing different oils, where distinct physical properties of each oil (e.g. viscosity) will affect delivery to the stock solutions. In addition, nominal concentrations do not account for changes in measured concentrations of hydrocarbons over time during the toxicity testing. The constituents of oil that partition from oil droplets to water are highly hydrophobic and may not remain in solution if they can adsorb to particulates, the surfaces of test vessels, etc. (Sprague 1969; Kiparissis et al. 2003).

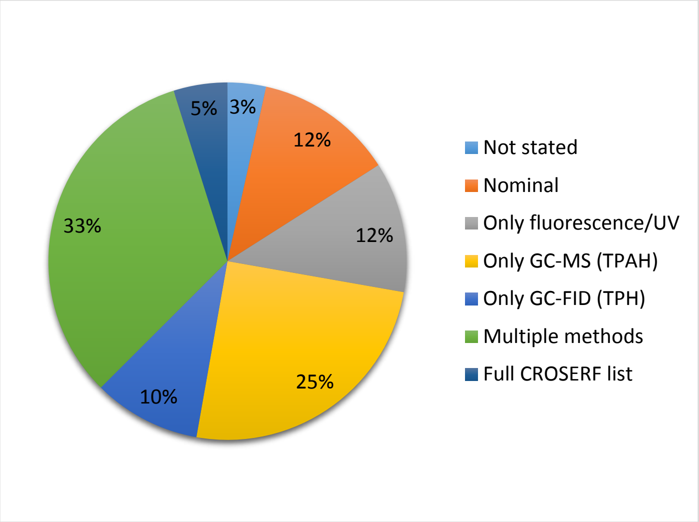

Figure 12: The proportion of the characterization methods used for test solutions in all papers reviewed (N = 144). "fluorescence/UV" is defined below. "Multiple methods" means at least one GC-based method combined with other GC or fluorescence/UV methods. There were 22 papers reporting volatiles (VOC), but this was always in combination with either TPAH or TPH analysis, so these are included in Multiple methods. "Full CROSERF list" indicates TPH, TPAH and VOC were done.

Chemical characterization of stock and test solutions to provide measured concentrations can overcome the limitations of using nominal concentrations (Aurand & Coehlo 2005; Forth et al. 2016), and can allow for comparison of toxicity test results among studies using different oils and even different exposure methods. Detailed chemical characterization can also broaden the use of the published data into other, likely unintended, applications and allows for detailed comparisons of toxicity among different studies (Redman & Parkerton 2015). Characterization of the oil and the test solutions can be used to quantify the change of specific targeted components of the oil solutions over time.

4.3.2 Analytical methods reported

The majority of the publications reviewed used at least one method to measure concentrations for at least one sample in the experiments (121 of 144; 84%; see Figure 13). Only 7 papers (5%) report chemical analysis covering the complete recommended list of components for the three groups listed under CROSERF. In practice, the analyses reported in each study varied due to a number of factors. Some choices were related to the objectives of the study; for example, when the toxicity of PAHs was of particular interest to a study, analysis of saturates (or alkanes) may have been omitted. Other choices were related to the oil being studied; highly weathered oils or heavy distilled oils are known to be free of volatiles including BTEX, and studies of these oils may have omitted VOC analysis. Still other choices may have been based on resources or practical considerations; a study with a minimal budget might choose only to do less expensive GC-FID analysis and forego the more expensive GC-MS analysis. The methods chosen will also be based on the design of the toxicity test. Although there have been recommendations to remove VOCs from the analytical requirements due to their volatility and low persistence (Barron & Ka'aihue 2003), *VOC analysis should be included to allow* *comparison among studies and oils, particularly for acute lethality tests.*

The scope of analysis is also important, and varies among studies. It may not be practical or feasible to analyze every sample prepared in a study with every available method; there will be a limit to the number of samples that can be analyzed over an exposure period. The entire dilution gradient should be sampled and analyzed at least once during a test to establish the relationship between dilutions and measured concentrations. A majority of publications (62%) measured oil concentrations in stock solutions (e.g., WAF or CEWAF), and at least one dilution of the stock or test solution, and reported the nominal and measured concentrations. When measured concentrations show good agreement with expected values based on transfer and dilution factors, it may be possible to estimate the concentrations in additional samples based on these factors (Forth et al. 2016). This can be particularly useful in studies with multiple dilutions where concentrations in the most diluted samples fall below the analytical detection limit.

The criteria for the number of samples collected and frequency of sampling is dependent on the experimental design (e.g., exposure regime, duration) and should be planned carefully to define the dynamic nature of test solutions. For example, in a static daily renewal exposure regime, samples should be taken within the 24-hour span to describe the time-varying exposure regime due to uptake of hydrocarbons by the fish and test vessel (Sprague 1969). Daily sampling of at least the highest concentration is also required to establish any longer-term changes in exposures, and to provide a measure of variance.

The inconsistency in the use of analytical methods is also problematic because of variations in the definitions of compound groups among studies and over time. Because oils are such complex mixtures, it is not possible to identify and assign every compound present in every oil to a designated group. Bejarano et al. (2013) pointed out the need to consider TPH definitions of C_7-C_{30}, C_8-C_{42}, and C_{10}-C_{36} in a review of other studies. The papers selected for this review also include TPH definitions of C_9-C_{44} (Gardiner et al. 2013) and C_9-C_{32} (Hemmer et al. 2011).

This has been further complicated by recent improvements in both capability and accessibility of GC-MS analysis, resulting in a constantly changing list of analytes to consider. Analysis of PAH and aromatic compounds in general is discussed in the next section. There is now a growing list of alkanes (mainly straight-chain or “n” alkanes) that can be identified and quantified by standard GC-MS measurements. Some studies are using these partially to fill the role that standard TPH analysis may have otherwise played, which is acceptable as long as a “sum of n-alkanes” term is clearly stated and not directly compared with a TPH number, unless it is validated to be equivalent to a TPH result.

Alternative measurements of oil or component groups without the use of chromatography are gaining in popularity. The most common are fluorescence and absorbance spectrometry (referred to as fluorescence and UV, respectively), with infrared spectrometry also reported. These take advantage of the optical properties of the aromatic constituents in oil, providing a means to detect the oil in many water solutions or extracts. These methods are rapid and low-cost compared with GC methods, and may have comparable or better detection limits. The main drawback of fluorescence and UV is that they are not specific to single or multiple compounds within the broader aromatic group, and a fluorescence or UV signal from one solution cannot be compared to a result from another solution unless additional information regarding the composition of the solutions is known. If a fluorescence or UV signal is carefully correlated with the components in a series of samples, usually through GC-MS analysis of a selected sample set, fluorescence or UV can be used to analyze related samples. For example, for replicate samples, for a dilution series, or over an exposure period, changes in the concentrations of detected components can be tracked on the assumption that the relative composition is reasonably conserved. Fluorescence spectroscopy has been used to measure the amount of oil in water based on standard curves relating fluorescence to known dilutions of the whole oil (e.g., Wu et al. 2012; Martin et al. 2014; Madison et al. 2015). This technique has also been utilized to measure dilbit dispersions and the fluorescence measurements of test solutions have been correlated with measurements of test solutions by GC-MS by Madison et al. (2015). In this review, 17 papers used fluorescence or UV as the sole method for determining test solution concentrations, which is better than nominal concentrations but of limited use. A further 16 papers used fluorescence or UV in combination with a GC-based method, which can provide additional information while reducing overall analysis costs and time.

Analyses selected for assessment of dilbit and solutions of dilbit in water should be consistent with methods for other oils. Depending on the experimental design, there may be an emphasis on different methods (e.g., more sampling for analysis of VOC in acute toxicity test of unweathered dilbit or more sampling for GC-MS in chronic toxicity tests of weathered dilbit). A shortcoming of the analysis of dilbit solutions in the published literature is lack of detailed measurement of test solutions at the end of experiments.

Priority Research Need

- Develop rapid and inexpensive analytical methods to chemically-characterize oil and test solutions (this includes high temperature distillation techniques).

4.3.3 Multiphase samples

The analysis methods specified in CROSERF and used in these studies do not distinguish between the dissolved and particulate phases in test solutions, primarily because water samples are extracted with solvents for analysis. Measurement of multiphase test solutions typically do not include protocols to separate the dissolved and particulate portions of the water samples, requiring careful review of the reported concentrations to avoid biasing the interpretation of toxicity tests (Redman et al. 2012). Toxicity of dissolved hydrocarbons will be underestimated if droplets inflate the measured concentrations of waterborne oil. The contribution of droplets in test solutions has been estimated based on comparisons of predicted aqueous concentrations (calculated from oil composition and solubility of individual components) to the measured test solutions (Redman et al. 2012), and solid phase extraction techniques have been proposed as a way of actually measuring the dissolved phase (Redman et al. 2014). The characterization of the particulate oil in the test solutions is also related to the stability and duration of the exposure. The exposure may be longer than the anticipated duration due to the presence of particulate oil. For example in a pulse or delayed effects exposure of fish embryos, the duration of exposure to oil will be extended or enhanced if oil droplets adhere to the chorion (Sørhus et al. 2015) and provide a source of exposure.

The presence of droplets can be demonstrated chemically by measuring poorly water soluble and highly persistent constituents of oil such as hopane, which is detectable in oil, but not as a dissolved hydrocarbon in water. Its presence in a solution is a clear indication of oil droplets. At the same time, chemical analysis comparing the WSF to the original oil has identified clear differences in chemical composition, confirming the presence of differentially-dissolved components in the water sample (Bennett et al. 1990; Forth et al. 2016). Maher (1982) used filtration with glass fiber to identify the contribution of particulate oil greater than the filter pore size in the WSF, but, filtration is not recommended due to the loss or change in dissolved fractions with handling (Girling et al. 1994; Singer et al. 2001a). Specifically, oil accumulating on filters can act as a source of hydrocarbons to water as it passes through the filter.

Characterization of oil droplet sizes in test solutions over time will contribute to the understanding of the variability among test solutions due to preparation methods, types of oil products, and weathering states.

Research Needs

- Assess the interactions of dissolved and particulate oil (physically- and chemically-dispersed) with dissolved and particulate organic matter, and their effects on the measured toxicity of oil.
- Develop inexpensive and practical methods for measuring the concentration and size distribution of oil droplets in test solutions to distinguish dissolved from particulate oil.
- Optimize the application of passive samplers for characterizing concentrations of dissolved hydrocarbons in test solutions.

4.3.4 Number of analytes measured by GC-MS

As the range of individual compounds in the PAH group that can be detected by GC-MS grows, the meaning of parameters like TPAH becomes increasingly problematic. The term TPAH creates a potential bias in the comparison of the measured concentrations in oil and test solutions if the number of analytes and their detection limits differ among studies. Recent advances in GC-MS methods in the wake of the Deepwater Horizon well blowout have expanded the PAH analyte list to over 50 compounds or compound groups (i.e. all aPAH isomers of a particular carbon number are summed to give a single result, e.g. C_3-phenanthrenes). Key additions to the original CROSERF list include benzothiophenes, naphthobenzothiophenes, and their alkylated homologs. A separate acronym ‘TPAH50’ has appeared to distinguish these TPAH measurements (Forth et al. 2016; Alloy et al. 2016). It is now common practice to report the total PAH concentration as the sum of 50 or more PAH analytes, though use of the TPAH50 acronym is not standardized (e.g., recent dilbit studies; Madison et al. 2015; Philibert et al. 2016). The ability to compare toxicity between past and future studies using a common metric of oil concentrations requires that publications report a complete list of analytes.

Of the 81 papers that measured TPAH, the most frequently measured number of analytes was 30-39 at 21% (Figure 13). Because the CROSERF list includes 40 PAH, the number of analyses in the 40-49 and 50-59 ranges (19% and 14%, respectively) is encouraging. All of the analyses with 50 or more TPAH compounds (24 papers) were published after 2007, reflecting recent improvements in GC-MS, although 22 other papers published after 2007 reported fewer than 40 TPAH analytes.

Research Need

- Develop data quality objectives to assess the reliability of chemical analyses of oil and of test solutions and to compare results and performance among multiple laboratories.

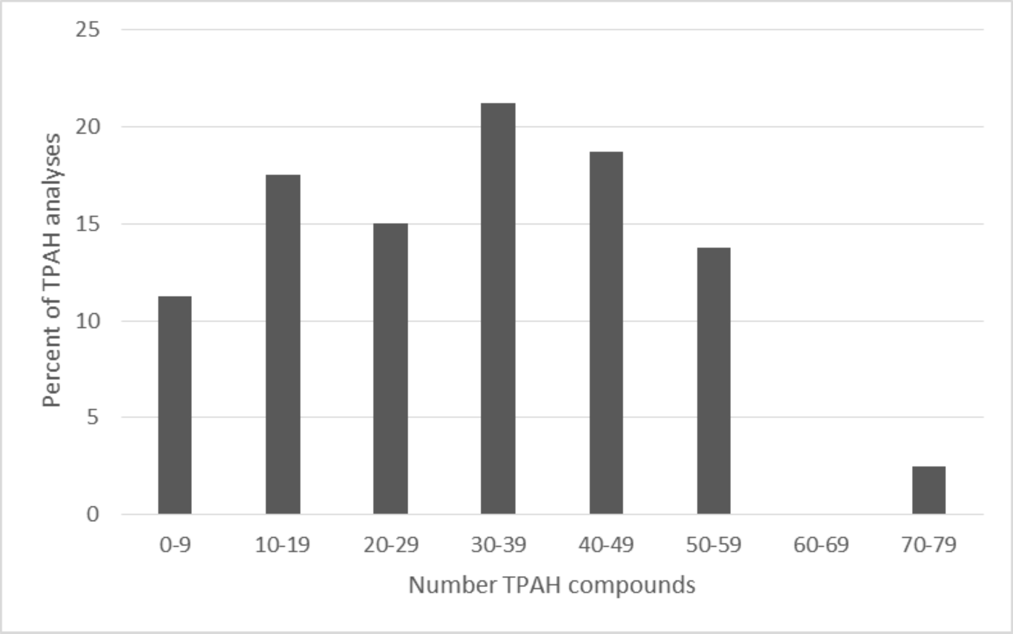

Figure 13: The proportion of analyses based on number of TPAH compounds reported. The CROSERF list has 40 TPAH compounds.

4.4 EXPOSURE REGIME AND OUTCOME

4.4.1 Species, life stage and duration of exposure of test organism

CROSERF recommended the use of temperate species. Recent oil exploration and development in the Arctic has expanded the array of relevant test species and test conditions. For example, Barron and Ka'aihue (2003) recommended the use of larval Pacific herring and Gardiner et al. (2013) assessed the toxicity of ANSC to copepods (*Calanus glacialis*), juvenile Arctic cod (*Borreogadus saida*), and larval sculpin (*Myoxocephalus* sp.) in conditions typical of open water of the Beaufort and Chukchi Seas (i.e., 0 to 5˚C, 20 to 35 ppt salinity). Nevertheless, the species selected for standard toxicity tests will most likely be those that are locally available or specified in test guidelines (e.g., rainbow trout, Environment Canada 1998).

A total of 74 fish species were captured in the review, 24 of which are species identified on the Department of Fisheries and Oceans list as aquatic species living in Canadian waters. The list of species captured were grouped as marine or freshwater and by cold or warm weather habitats in Appendix 3, and species from Canadian waters are underlined. Test species used in 10 or more studies were Pacific herring (10), pink salmon (11), rainbow trout (11), and inland silversides (17).

Approximately 15% of the studies tested more than one species. Of these, several used multiple methods with multiple oils (e.g., Swedmark et al. 1973; Anderson et al. 1974; Hedtke & Puglisi 1982; Woodward et al. 1987; Fucik et al. 1995; Clark et al. 2001; Lee et al. 2011; Adams et al. 2014b). Many compared toxicity to fish with toxicity to other aquatic organisms (e.g., crustaceans, benthic invertebrates), but the scope of this review was restricted to fish.

The duration of the exposure is the time (days) that organisms were exposed to oil. Exposure durations were categorized for each test reviewed, resulting in 199 total entries (Figure 14). If the exposure duration was not stated explicitly, it was categorized as not specified (**NS**). Exposure durations were also classified as **Acute** (≤ 4 days; usually lethality tests with juvenile or embryonic life stages)**, Chronic** (>4 days; mostly embryos or larvae), or **Both** if a study included both acute and chronic exposures Figure 15). Delayed effects that were expressed beyond the period of the oil exposure were not coded, only the duration of actual oil exposure.

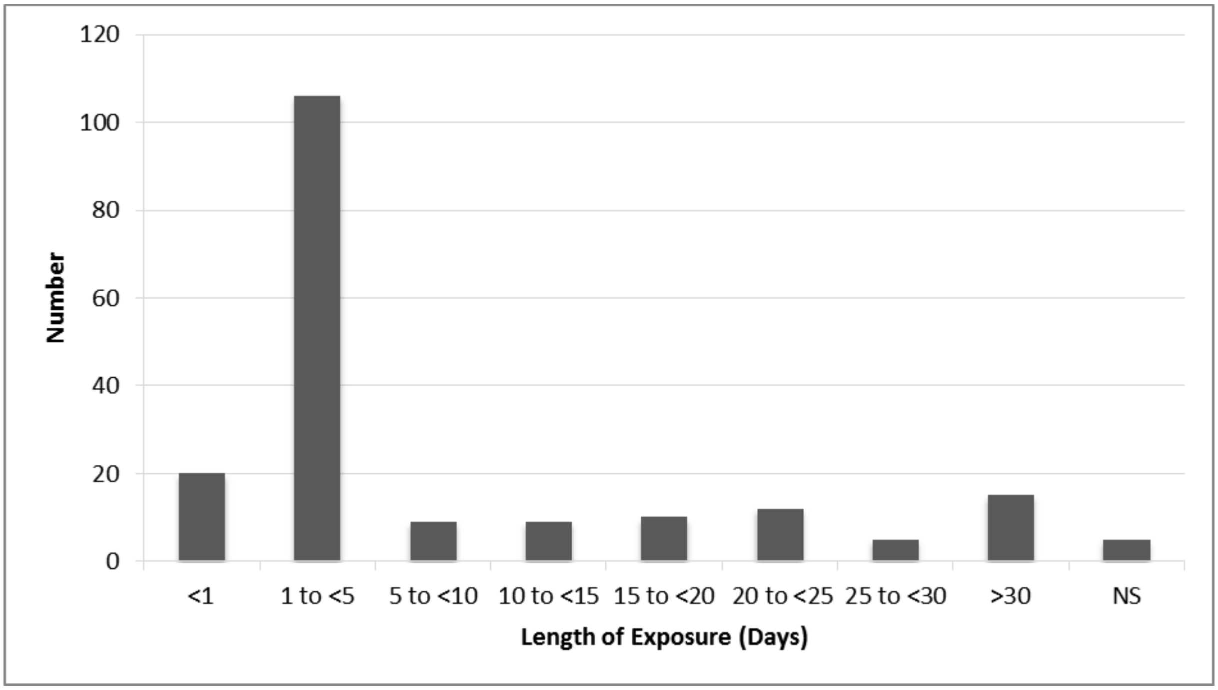

*Figure 14: The number of exposure durations (days) for toxicity tests. If the exposure duration was not explicitly stated, the publication was classified as not specified (**NS**). Data with ranges were not included (N=8).*

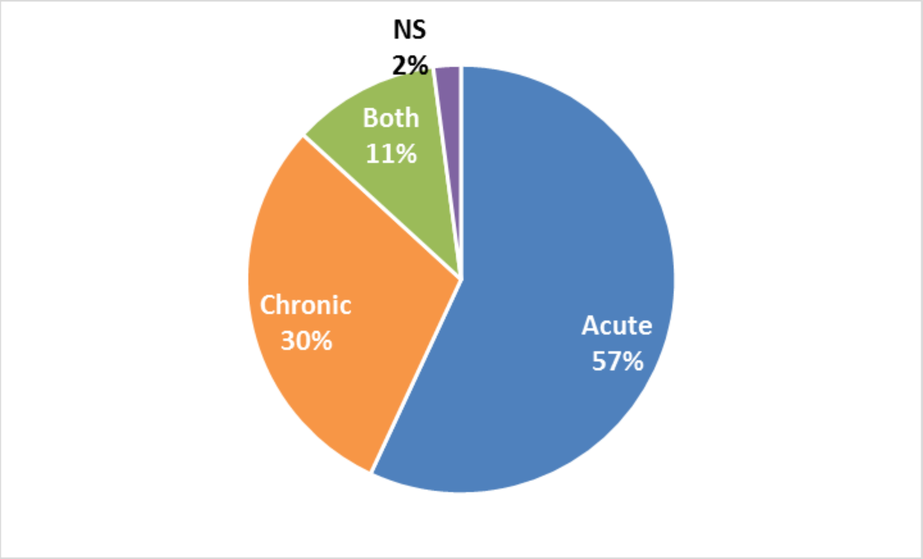

*Figure 15: The percentage of toxicity test exposures classified as **Acute** (≤ four days) or **Chronic** (> four days). Publications including both acute and chronic exposures were classified as **Both**. If the exposure duration was not given, it was classified as not specified (**NS**).*

Acute exposures were more common than chronic and few studies tested both. The full range of exposure durations was 0.04 to 240 days. The most common was 1 to <5 days (53.3%), with four days (96 hours) the most frequent. Four-day (96 h) acute toxicity tests are the standard recommended by CROSERF (Aurand & Coelho 2005). In some toxicity tests, exposure times for acute lethality tests were shorter than 96 h because experimental designs simulated estimated exposure times during actual spills, but LC50s for brief exposures confound comparisons of toxicity among oils. *To avoid this problem, acute lethality tests should be run for at least 96 hours, with observations of mortality at intervals of 1, 2, 4, 8, 16, 32, 64 and 96 h to provide LC50s for both short and standard exposure times (Sprague, 1969).*

In addition to characterizing exposure durations as acute or chronic, each test species was classified by life stage (Figure 16). As some species were tested at more than one life stage in a single study, there were 165 entries. The most frequently tested life stages were embryos (31.5%), juveniles (26.1%), and larvae (3%); a further 9.1% of papers did not specify life stage.

Juvenile and adult fish were typically used when the purpose of the test was to assess acute lethality or for tests of physiological responses. Chronic tests were typically conducted with embryos to assess impacts on reproduction. Sensitivity to oil exposure varies with life stage of the test organism and the nature of effects; the early developmental stages of fish tend to be most sensitive (McIntosh et al. 2010; Binder & Stageman 1983). Barron and Ka'aihue (2003) recommended two exposure periods for each species, eight hours to assess acute responses and seven days to capture delayed effects, but not mimic chronic exposures. Delayed effects could also be assessed by holding fish in clean water following oil exposure. Exposure durations longer than 30 days have been recommended to best represent all of the effects of an oil spill, including delayed effects (Heintz et al. 2000).

Experiments conducted with well-characterized species, oils and endpoints provide important bench-marks for assessments of dilbit toxicity.

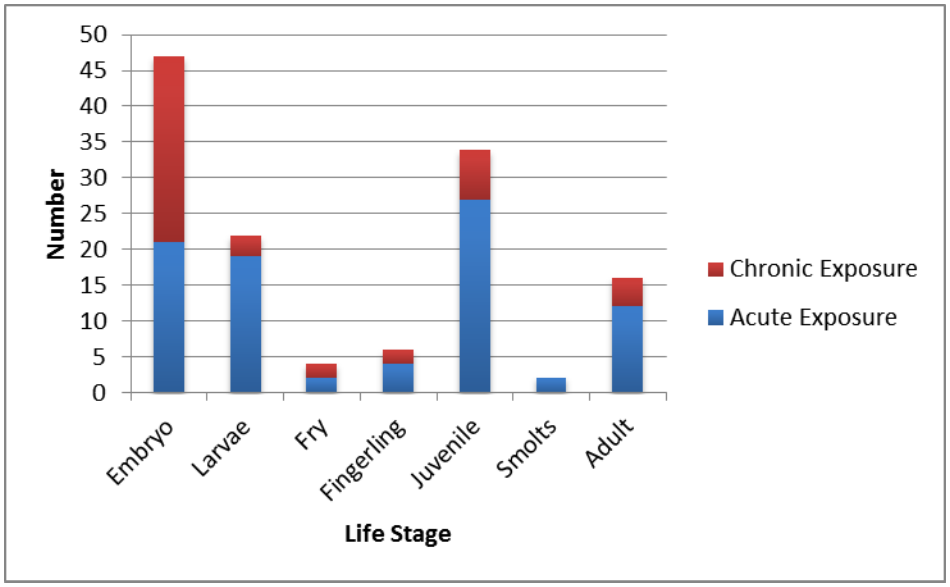

*Figure 16: The number of tests with different life stages of fish, differentiated by acute (≤4 days) or chronic (>4 days). If the exposure duration was not stated explicitly, it was classified as not specified (**NS**).*

Research Need

- Toxicity tests should be repeated with various dilbit products and different species typical of Canadian environments, using experimental designs that adhere to standardized methods.

4.4.2 Abiotic factors of fish exposure

Some of the abiotic factors reported in toxicity test methods were photoperiod and the characteristics of dilution water, including pH, temperature, conductivity, alkalinity, hardness, and dissolved oxygen. Many of these characteristics were not reported in detail in most publications. When reported, they were either recorded as averages for the exposure period or measurements were provided at specific times during the exposure period.

The necessity to measure dissolved oxygen in test solutions was highlighted for tests with industrial effluents (Doudoroff et al. 1951). In exposures with sealed vessels without aeration, oxygen may become depleted, even with daily static renewals of test solutions (CONCAWE 1993, Girling et al. 1994). The CROSERF method recommended that dead animals not be removed from the test solutions during the exposure. However, in tests at temperatures greater than 10˚C and longer than 24 hours, decay of dead fish could decrease dissolved oxygen concentrations (Sprague 1995) and rates of microbial degradation of oil would likely increase. In accordance with technical guidance documents for fish toxicity testing (e.g., ECCC, 1998; 2007; 2011), all dead organisms should be recorded and removed immediately when found.

Most toxicity reports (63.4%) did not specify the photoperiod for the test. Where described, photoperiods most commonly reported included 0:24, 12:12, 14:10, and 16:8 hours light:dark. Photoperiods are often determined by the acclimation and holding conditions specific to the test species (e.g., ECCC 1998; 2007; 2011) rather than any consideration of the interaction of oil with light. The guidance from ECCC (1998) and ECCC (2007) for photoperiods is specific to the species and life stage. A 16:8 light:dark photoperiod with natural or fluorescent lighting was recommended for the entire period for early life stages and for fingerlings of rainbow trout.

These same protocols recommended that light regimes should be changed when working with chemicals that are photoactivated or degraded with UV, a recommendation that is relevant to tests with oil. The potential for photoenhanced toxicity is related to the lighting source (UV spectrum), exposure regime, species, and life stage. A detailed review of lighting options can be found in ASTM E1733-95 (2014), *Standard Guide for Use of Lighting in Laboratory testing*. Lighting sources include solar lamps (natural solar spectrum is UVA (315-400 nm) with the remaining UVB (280-315 nm at 5%), fluorescent lamps (standard laboratory lighting with minimal UV (Barron et al. 2005) or darkness (Lee 2003). The concern is that the light regime during testing may change the composition and toxicity of the test solution through photodegradation or photomodification. The degree of photooxidation is PAH-specific and dependent on the wavelength spectrum of the light (Lee 2003). Irradiation of WAF with UV light prior to toxicity tests with juvenile silversides did not increase the toxicity (Little et al. 2000). Similarly, exposure of WAF or CEWAF of weathered ANSC to sunlight prior to tests with herring larvae did not change their toxicity. In contrast, the toxicity of weathered ANSC WAF and CEWAF to herring larvae increased by up to 50-fold when larvae were exposed to sunlight during the toxicity test (Barron et al. 2003).

The mechanism of this dramatic photoenhanced toxicity is phototransformation of PAH accumulated in semi-transparent tissues (Barron & Ka'aihue 2001) of eggs and larvae (e.g., Barron et al. 2003) and juvenile fish (Little et al. 2000). Barron and Ka'aihue (2001) noted the potential for up to 1000-fold increased toxicity of oil and PAH following exposure to daylight levels of UV compared to exposures to typical fluorescent lighting in laboratory tests. Similarly, the mortality of juvenile silversides exposed to WAF of weathered crude oil was 5.6-fold higher in UV exposures than in exposures mimicking office lighting (Little et al. 2000). The potential for photoenhanced toxicity is dependent on the species and life stage; translucent test organisms are at higher risk of photoenhanced toxicity than those that have well-developed pigment in the skin (Barron et al. 2005).

The UV spectra of laboratory lighting are rarely reported in oil toxicity studies. Although UV wavelengths from laboratory lighting do not appear to be problematic, it is recommended that toxicity tests be conducted with low intensity laboratory lighting that includes a minimum of light energy at UV wavelengths. Measurement of UV intensity in toxicity tests has been recommended for standard testing as it may be a useful component in models and can be directly linked to site-specific conditions (Barron & Ka'aihue 2001; 2003; Bragin et al. 2016). *Alternatively, the potential bias of photosensitization could be identified by adding a UV control in which fish are exposed to one or more concentrations of oil in the dark.*

Research Need

- Assess the extent to which dilbit products are photosensitized in laboratory tests and under the range of UV intensities typical of Canadian environments.

4.4.3 Size of exposure vessel

The size of exposure vessels, or more properly, the volume of test solutions, plays a key role in determining the extent of exposure of test organisms and whether test conditions are constant. Based on a review of respiration rates of fish, Sprague (1969) recommended that the volume of test solution at each test concentration be sufficient to meet the oxygen requirements for the test species. The purpose was to ensure that respiratory stress was not a confounding factor in toxicity tests without aeration. This criterion was also a surrogate for ensuring that test solution volumes were sufficiently large that bioaccumulation of the test substance by fish did not markedly reduce test concentrations, assuming that rates of oxygen uptake were roughly equivalent to rates of toxicant uptake.

For juvenile rainbow trout, the recommended volumes were equivalent to 2 to 3 L of water per gram of fish biomass per 24 h. The current Environment and Climate Change Canada standard test method for acute lethality of single substances or effluents (EC 2007) uses a similar criterion but expresses it as a maximum loading rate of < 0.5 g fish biomass/L at 15˚C (i.e., 2 L/g/d). Following this recommendation, a four day static non-renewal test with ten 1.0-gram fish would require 80 L of test solution (a large aquarium). If solutions are renewed every 24 h, solution volumes could be correspondingly lower (20 L) although the total volume over 4 days would remain at 80 L. For flow-through tests, much smaller test containers could be used, provided that the total flow of test solution was equivalent to 80 L over 96 h (i.e., a minimum flow of 14 mL/min). For tests with 10 fish embryos weighing 1 mg each, the recommended volume of test solution would be 1000-fold smaller, i.e., 80 mL over 96 h.

The biomass loading rates and test volumes recommended by Sprague (1969) were developed for trout, and should be adjusted for different species and different temperatures based on reported or measured rates of oxygen consumption. If these data are not available, the requirements for trout are conservative and could be adopted for most species.

While the recommended biomass loading rates for toxicity tests should limit the depletion of test substances in solution, there is an additional problem for hydrophobic hydrocarbons in tests with oil. Hydrophobic compounds partition from water to surfaces of tanks. As a consequence, concentrations in solution will decrease quickly over time, even in the absence of fish. With decreasing size of test containers, there is an increasing surface area-to-volume ratio, which will accelerate the proportional loss of hydrocarbons from solution. While the rate of partitioning will remain constant, the proportional loss of hydrocarbons from small containers of test solution will be much greater than from large containers. Independent of fish size, larger tank volumes will provide a greater stability of test solutions and a longer half-life of specific components. The loss of hydrocarbons from test solutions would be further aggravated by particulate or dissolved organic material derived from fish wastes or uneaten fish food. For this reason, feeding is not recommended prior to, or during, acute lethality tests (i.e., 96 h) (EC 2007). *Where feeding is required to maintain the health of fish during more prolonged exposures, extra effort is needed to replace test solutions and test containers immediately after feeding, and to clean test containers daily to avoid the accumulation of debris and biofilms that absorb hydrocarbons or harbour bacteria that can create metabolites of hydrocarbons.*

The detailed review of published data on oil toxicity indicated that most papers (65.1%) reported tank or test solution volumes, which ranged from <1 mL (embryos in microplate cells) to <100 L (juvenile fish in tanks) (Figure 17); the most common vessel size range was 10 to < 100 L (27.6%). It is difficult to put these volumes in context, because total loading (g biomass/L/d) of fish in each tank was not often reported. While the very small test vessels (≤100 mL; 8.6% of published tests; Figure 17) may have been convenient and economical, the estimated toxicities were likely confounded by changes in oil concentrations during the test due to excess biomass or absorption to tank surfaces. Overall, the potential bias is to underestimate toxicity. The only way to ensure that tank volumes have not confounded toxicity tests is to measure the concentrations of test substances in test solutions and to calculate toxicity on the basis of measured values (see section 4.3 on chemistry). *Because of the importance of this bias, test tank volumes and the biomass of fish in each tank should always be reported*.

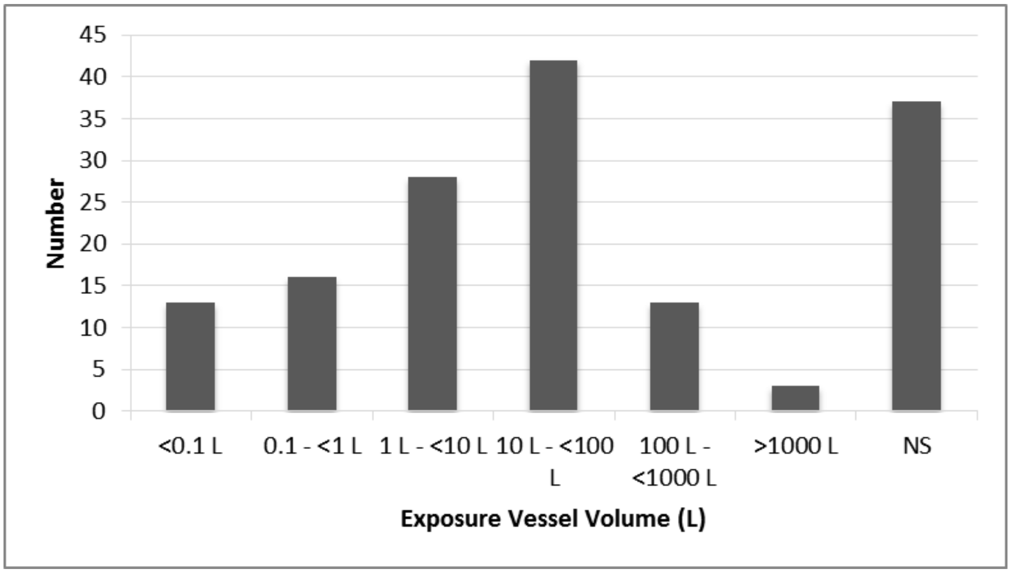

*Figure 17: The number of exposure vessel volumes (L) used in oil toxicity tests. If the exposure vessel volume was not explicitly stated, it was classified as not specified (**NS**) (N=152).*

Because dilbit weathers quickly and is more adhesive than conventional crude oil, the potential losses of dilbit to exposure vessels should be considered in experimental designs and the extent of losses measured by chemical characterization of test solutions and accumulation of oil on vessel walls over time.

4.4.4 Material of exposure vessel

Exposure tanks may be constructed of glass, stainless steel, or different forms of plastic (e.g., polyethylene, polycarbonate, Teflon), and the same materials may be used in tubing, pipe fittings, or devices used to prepare test solutions (pipettes, beakers, stir bars, etc.). Most petroleum hydrocarbons are hydrophobic, and will partition from test solutions to materials that comprise any part of a test system. Similarly, oil droplets in suspension will adhere to most surfaces. As a result, hydrocarbon concentrations in test solutions decline during solution preparation and toxicity tests (e.g., Wu et al. 2012), and may approach non-detectable within 24 h, the typical solution replacement interval for static renewal assays. Intuitively, plastics that absorb organic materials (e.g., low-density polyethylene) should be the least suitable for oil toxicity tests, and would cause the largest and most rapid changes in oil concentrations. Alternatively, glass and stainless steel should absorb the least amount of dissolved hydrocarbons and be the materials of choice, although oil droplets may still coat these surfaces. Tests of the bioavailability to fish of hydrophobic compounds from pulp mill effluents demonstrated that glass aquaria constructed with silicone seal retained measurable concentrations of hydrophobic compounds following a test. Sufficient partitioned from the contaminated silicone into un-contaminated water to cause a measurable accumulation by control fish (Hodson, unpublished observations). The capacity of silicone seal to accumulate hydrophobic compounds is the basis for recommendations to use silicone films as passive samplers to chemically-characterize test solutions (Redman et al. 2014). *Therefore, oil toxicity tests should not be conducted with aquaria constructed with silicone seal because the results will be biased by a faster loss of hydrocarbons from solution than in all-glass vessels, and there will be an increased risk of contaminating control solutions when tanks are re-used, reducing the capacity to detect toxicity.*

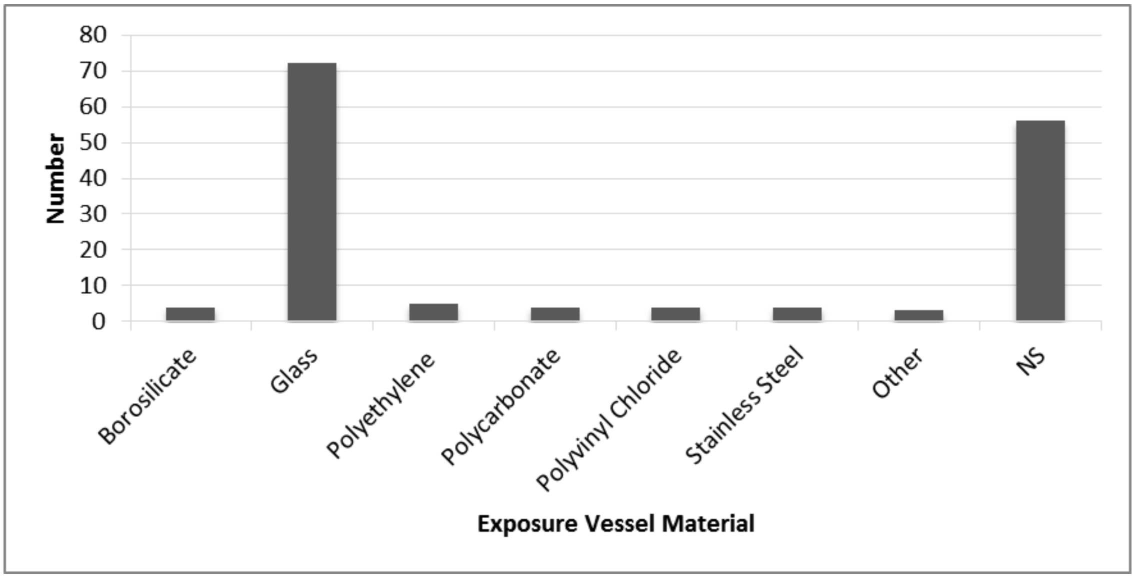

Figure 18: The number of exposure vessel materials reported in this literature review (N=152). Materials that were reported ≤ three times were classified as ***Other****. If the material was not stated explicitly, it was classified as not specified (****NS****).*

Of 144 studies reviewed, 152 exposure vessel entries were identified. Of the 152 entries, 63.2% reported the type of material in test tanks (Figure 18), of which 80% was glass. If an exposure vessel material was reported three or fewer times, it was categorized as **Other**; these entries included concrete and fibreglass exposure vessels. The remaining 36.8% did not specify the type of material of exposure tanks, and were designated as not specified (**NS**). No study provided specific descriptions of tank construction and the potential for contact of silicone with test solutions. There were no systematic assessments of the interactions between test tank materials and the extent and nature of changes in solution composition; this is an important research need, particularly for viscous oils rich in resins and asphaltenes, such as dilbit products.

Research Need

- To identify the most suitable toxicity test vessels, assess the absorption of oil from test solutions to various tank materials and the extent, time-course, and nature of changes in the chemical composition of test solutions.

4.4.5 Exposure regime

Test organisms can be exposed to test solutions in static exposures, semi-static exposures, pulse exposures or continuous flow systems. Static non-renewal (**SNR**) exposures refer to tests in which fish are exposed to a single aliquot of test solution for the duration of the test, with no renewal or refreshment of the test substance, often for periods up to 96 h. If the test substance is highly water soluble (e.g., salt), the exposure concentrations may remain constant throughout the test. Where the test substance is hydrophobic (e.g., oil, PAH), there is a continuous decay in concentrations throughout the test, often to undetectable concentrations within 24 h of the start of the test (Kiparissis et al. 2003). Semi-static or static renewal (**SR**) tests provide a periodic renewal of test solutions, which remain static for defined intervals. For example a static daily renewal protocol would involve solution replacements every 24 hours. For substances that decline in concentration throughout the test, the exposure regime resembles a series of spikes, often decaying to zero within the renewal interval. Pulse exposures (**P**) alternate exposure to test solutions with exposure to clean water, with fixed or variable durations of each (Aurand & Coelho 2005). Some may include a single period of exposure followed by a prolonged exposure to clean water to assess the potential for delayed effects. Constant flow-through protocols, (**CFT**) renew test solutions by means of a continuous addition of test substance to a continuous flow of water, with little or no decay in concentrations over time, although there may be random variation (e.g., Nordtug et al. 2011). Constant exposures (**C**) can also be achieved through partition-controlled delivery, in which test substances partition to water from high concentrations in or on a solid substrate (e.g., Kiparissis et al. 2013). Each test concentration is determined by the concentration gradient between the film and water, the partition coefficient of the test substance, and its water solubility. For tests with oil, a common strategy is to pass water through a column filled with gravel coated with oil (Marty et al. 1997). While this gives more stability of concentrations in acute exposures (< 1 week), concentrations ultimately decay significantly over weeks and months in chronic exposures as the films of oil are depleted of the more water soluble hydrocarbons by weathering (water washing) (Martin et al. 2014).

The static renewal, static non-renewal and constant flow-through regimes were the most frequently used, accounting for 89.8% of all tests (Table 6).

Exposure protocols that generate non-steady-state concentrations may mimic site-specific conditions at a spill. For example, the CROSERF method of testing oil involves the continuous addition of clean water to each concentration of oil in a gradient, so that concentrations decline to non-detectable within 6 hours or less. While this might represent the rapid dilution of spilled oil at an open-water site, it would not represent the conditions at a well blow-out or a pipeline spill where large amounts of oil are discharged over a prolonged period.

*Table 6: Number of exposure regimes. Exposure regimes were classified into semi-static or static renewal (**SR**), static non-renewal (**SNR**), constant flow-through protocols (**CFT**), pulse exposures (**P**), and constant exposures (**C**). When exposure regimes were not specified, the entry was classified as **NS**.*

Exposure Type	Number	% of Total Entries
SR	45	26.9
SNR	61	36.5
CFT	44	26.3
P	4	2.4
C	10	6.0
NS	3	1.8
Total number of entries:	167	N/A

The results of tests with time-varying concentrations are difficult to characterize, because end-points (e.g., 96 h LC50s) could be based on the initial concentration, the final concentration, the average concentration, or the time-integrated concentration (Lee et al. 2015). The total exposure is also important, not just the peak concentration. For instance, a continuous exposure of Atlantic haddock embryos to low concentrations of oil was as toxic as a pulsed exposure to a higher concentration (Sørhus et al. 2015). The different exposure regimes and preparation of oil-water solutions dictate the stability of exposure concentrations over time. The most stable exposures are those generated by a continuous production of fresh solutions of oil in water, (e.g., Nordtug et al. 2011). Continuous-flow systems and oiled-gravel columns likely provide the best exposure tools for true comparisons of toxicity among oils, among species, and among test conditions (e.g., temperature). Comparisons using other methods are much more likely to be confounded by differences among labs in test conditions, or the interactions of the characteristics of the oil with the test method. For example, preparations of WAF of light oils by stirring are likely to incorporate more fine particulate oil and dissolved hydrocarbons into solutions than WAFs of heavy oils, primarily because differences in the viscosity and interfacial tensions between heavy and light oils determine the extent to which droplets are formed for a given level of mixing energy.

Comparisons of toxicity among oils can also be improved by chemical dispersion (CEWAF). Dispersants interact with films of oil to reduce both the effects of viscosity on droplet formation and the confounding effects of oil viscosity on apparent toxicity. In comparisons of bioavailability or chronic toxicity among oils, the amount of CEWAF added to test solutions to cause a given level of effect was 6 to 1100-fold less than the amount of WAF needed (Ramachandran et al. 2004; Adams et al. 2014b). When the results were expressed as the amount of hydrocarbons measured in test solutions, there was virtually no difference in effects between WAF and CEWAF. This reinforces the need to measure the amounts of hydrocarbon in test solutions to enable toxicity comparisons among oils without the confounding effects of the method of solution preparation.

The choice of exposure regimes is critical in determining the outcome of toxicity tests and their utility in ERAs. In the past, a large proportion of oil toxicity tests applied static or semi-static exposures because of their simplicity and low cost, but if the results are not easily compared or applied in ERAs, the resources expended may be wasted. Research is needed to improve on existing test designs. For tests in which there are time-varying concentrations of oil in test solutions, the interactions among time, concentration, and extent of effects should be assessed with statistically-based models to determine if there are consistent and predictable relationships that would improve comparisons among test conditions and among test labs. *Designs of tests with time-varying concentrations of oil should be guided by observed temporal trends of oil concentrations at sites of actual oil spills. No matter which toxicity test method is selected for dilbit, in particular for unweathered products, chemical characterization of test solutions is essential to describe the change in the composition and concentration of oil over time.*

Research Needs

- Identify and evaluate practical and cost-effective test protocols to maintain constant concentrations of oil in test solutions over the duration of toxicity tests.
- Develop data quality objectives to indicate the reliability of analysis results for characterizing oils and solutions and to compare results among multiple laboratories.
- Develop statistical models that integrate oil concentrations in test solutions over time to generate a stable metric of exposure and calculated toxicity.
- Model toxicity over exposure time for oil toxicity tests with time-varying concentrations of oil to identify the best indicator of toxicity, i.e., the initial, average, final or time-integrated concentration of oil.

4.4.6 Effects measured and endpoints

The relationship between the severity of toxic effects and exposure to oil is used to calculate specific endpoints that can be compared between studies. Effects measured include those of lethal and sublethal consequences, such as mortality, development, and morbidity. Lethal and sublethal effects can be measured mid-exposure, directly after exposure, or measured at time point(s) following the exposure, after transfer to clean water (i.e., as prolonged or delayed effects).

Depending on the experimental design and the techniques for measurements, endpoints may be measured throughout the exposure duration or require sampling or analysis of the organisms at a specific exposure time or stage of development. Endpoints are expressed as the effect measured after a fixed duration of exposure. For example, the concentration of oil required to cause chronic toxicity to 50% of test organisms after a 7 day exposure, would be termed the 7 d EC50 or median effective concentration.

The no observable effects concentration (NOEC) is the highest concentration that does not cause a statistically-significant effect in comparison with controls. The lowest observable effect concentration (LOEC) is the lowest concertation that causes a response significantly different from controls. Effects can also be reported at specific effects levels. The NOEC/LOEC were identified as the most commonly used endpoints in ecotoxicology by Harris et al. (2014), but they have been widely criticized because of potential bias caused by uneven spacing of measured exposure concentrations and high sensitivity to variance at the threshold of toxicity.

Thirty of the 144 publications reported multiple endpoints. If the time at which the measurement was made was not indicated, then the endpoint was recorded without the time indicated. If a given endpoint was reported three or fewer times, it was classified as **Other**, these entries included EC10, EC20, EC80, LC10, LC90, and LC100. **NE** was used to indicate tests that did not calculate and endpoint value.

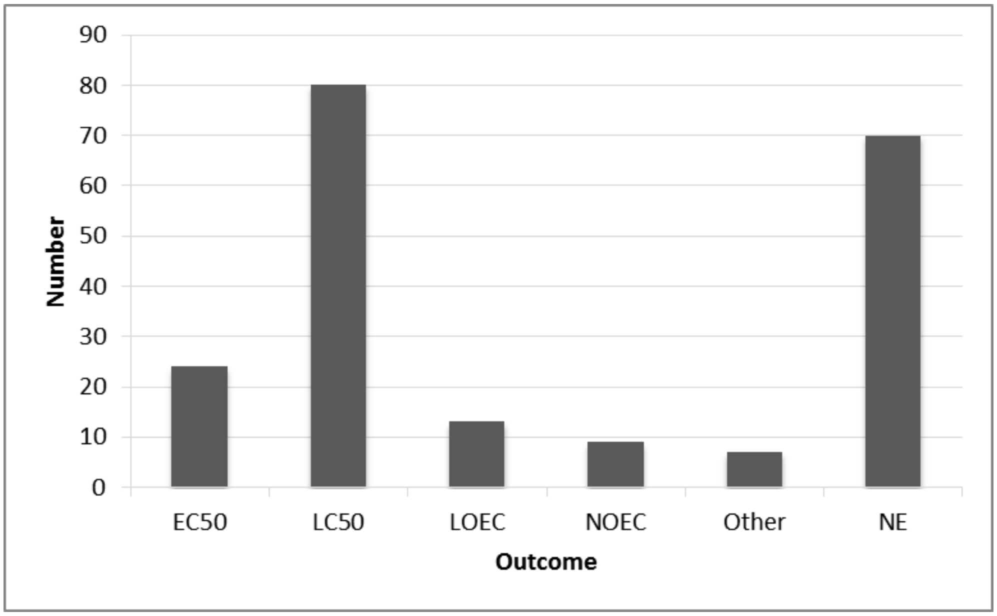

*Figure 19: The number of reported endpoints; median effective concentration (**EC50**), median lethal concentration (**LC50**), lowest observed effect concentration (**LOEC**), and no observed effect concentration (**NOEC**) (N=203). If a given endpoint was reported ≤ three times, it was classified under **Other. NE** was used to indicate tests that did not calculate an endpoint.*

For oil toxicity tests, **LC50s** and **EC50s** were the most common end-points reported. Sublethal effects were observed primarily in fish embryos and included malformations, stunted development, reduced swimming ability, blue sac disease (BSD), altered gene expression, and cardiovascular defects. BSD is a non-infectious disease characterized by lipid membrane breakdown and accumulation of fluid in yolk and pericardial sacs, body and ocular hemorrhaging, fin erosion, craniofacial malformations, and spinal curvature (Bauder et al. 2005). Fish with the most severe signs of BSD are not expected to survive to adulthood. Exposure to PAHs partitioned from oil is also commonly accompanied by altered gene expression, particularly genes that code for cytochrome P450 (e.g., CYP1A) enzymes (Timme-Laragy et al. 2007). Induction of CYP1A enzymes and enzymes mediating oxidative stress due to PAH metabolism are sensitive indicators of PAH exposure in aquatic organisms (Fragoso 1998, Madison et al. 2015). The endpoints for chronic sublethal effects (e.g., 17 d EC50) generally occur at lower oil concentrations than acute lethality, and are a more sensitive indicator of potential effects on fish populations.

Cardiovascular defects caused by exposure to petrogenic PAH (Incardona et al. 2009, Incardona et al. 2013) can lead to reductions in cardiac output and the swimming capacity of surviving juveniles and translate to lower survival rates of fish exposed in the wild (Incardona et al. 2009). With advances in molecular techniques, endpoints that are more sensitive have been developed to assess changes in physiological health, such as gene expression in fish embryos exposed to dilbit (e.g., Madison et al. 2015).

4.5 OVERVIEW OF REPORTED OIL TOXICITY TEST METHODS

4.5.1 Variability in methods captured in review

This review captured a variety of methods which potentially change the outcome of toxicity tests. This is a partial review that does not include a diversity of research, particularly grey literature, outside the scope of the review.

The literature published after the development of the CROSERF guidance documents indicates that the variation in oils tested, methods used to generate test solutions, and methods to expose fish to oil may still lead to misinterpretation of results and inhibit comparisons among oil toxicity studies.

The variability among methods for toxicity testing of WAF was greater than that for CEWAF. This may be related to the apparent lack of an organized effort to standardize WAF methods for other environmental scenarios, such as oil interactions with bed sediments or the high energy mixing of oil in rivers. This has prompted several research groups to develop their own methods, including methods to generate HEWAF solutions (e.g., Alloy et al. 2016).

The diversity of methods is related in part to the different aims of tests (e.g., research versus regulatory testing), modifications to standardized protocols to suit different scenarios, and development of alternative methods. George-Ares and Clark (2000) found variability in dispersant toxicity based on data compiled from the literature, because not all studies were conducted to support decision making for dispersant use.

4.5.2 Review of methods used to assess the toxicity of dilbit

There have been three studies of dilbit toxicity to fish published in the peer-reviewed literature, Madison et al. (2015), Philibert et al. (2016), and Alderman et al. (2016). The first two referenced a modified CROSERF protocol described by Singer et al. (2000) as the basis of their design to generate test solutions.

Madison et al. (2015) assessed the toxicity of AWB WAF and CEWAF to Japanese medaka embryos from fertilization to hatch (17 day exposure). The methods to prepare WAF and CEWAF test solutions were based on Adams et al. (2014a,b) and Martin et al. (2014) who applied the most common modifications of the Singer et al. (2000) protocol. In short, solutions were prepared daily in graduated beakers with low energy mixing of 1:9 oil-to-water ratio (OWR) on a magnetic stir plate (20-25% vortex depth). WAFs were mixed for 18 hours with a 1 hour settling period; CEWAFs were mixed for 18 hours, at which time Corexit 9500 was applied in a 1:10 DOR for an additional 1 hour of mixing, and 1 hour of settling time. In both WAF and CEWAF, the floating oil was removed from the beaker to enable collection of the aqueous phase. The authors noted variability in the efficiency of dilbit dispersion and speculated that difficulties in dispersing dilbit may have confounded the measured toxicity.

In Philibert et al. (2016), the toxicity of dilbit WAF to zebrafish embryos was compared to WAFs of medium sour composite (MSC) and mixed sweet blend (MSB). In this experiment, WAFs were also prepared with methods based on Singer et al. (2000). Oil and water in a 1:10 ratio (1.8 L embryo media plus 180 mL oil) were mixed in a closed 2 L aspirator bottle at approximately 100 rpm for 20 hours, and left to settle for 4 hours, after which the unfiltered WAF was collected, pH-corrected, and diluted into test solutions for daily static exposure of zebrafish embryos from 30 minutes to 7 days post fertilization.

Alderman et al. (2016) exposed sockeye salmon parr to WSF generated by a continuous flow of water over ceramic beads coated with CLB summer blend for one or four weeks. Beads were soaked with dilbit for approximately 30 minutes. Water passed over the different loadings of beads and was collected in 2 L header tanks to be pumped into two tanks for the experiments.

Our experience (Adams, unpublished observations) with preparing AWB and CLB winter blend CEWAF with medium energy mixing by magnetic stir plate was that the depth of the oil vortex increased during the first few hours of the mixing period, which was not observed with CEWAF of conventional crude oils. This meant that the mixing energy required to create a 20-25% vortex depth was greater during initial mixing than a few hours later when the mixing energy was reduced to maintain the 20-25% vortex and to prevent the vortex from hitting the stir bar. It may be that the stir plate did not sustain a constant mixing speed, but equally, the stirring speed could change with the change in the properties of the oil being stirred. This change in the behaviour of the stirred oil:water system may be attributed to weathering of the dilbit during 18 hours of stirring, and likely contributed to the variability among CEWAF preparations for the same oil. We also noted that the time to the breaking of the vortex after application of dispersant varied from day to day. Methods with a higher mixing energy may generate more repeatable dispersions by overcoming the early resistance to forming a vortex.

More detailed studies are needed to understand the changes in physical and chemical properties of dilbit during preparation of test solutions, and during the toxicity test itself, and how those changes interact with observed toxicity.

5 PROPOSED FRAMEWORK TO DEVELOP STANDARD PROTOCOLS

This literature review demonstrated a diversity of methods for all aspects of toxicity testing, with many variations on standard protocols that were often poorly described. For the CROSERF protocol, there have also been several reviews that identified protocol limitations and that suggested modifications to methods such as mixing oil and water or applying dispersant (section 3.3.4). Therefore, it is difficult to recommend specific methods, or changes to methods, to respond to the unique characteristics of dilbit without first addressing the issues associated with oil testing in general. Recommendations to deal with any oil that has unique properties (e.g., related to testing of highly weathered oil, Barron & Ka'aihue (2003) section 4.3) may eliminate the need for methods unique to dilbit. Protocols must be sufficiently flexible to match the diversity of objectives inherent in regulatory toxicity testing, ERAs and EIAs, and to generate data that are readily compared among studies. Given that the original CROSERF process was consensus-based, involving practitioners from industry, government and academia, a similar process should be followed to ensure that any recommendations will meet the needs of those who use the methods. The proposed framework presented in this chapter outlines the CROSERF protocol for all aspects of toxicity testing (Appendix 1), updated to recognize widely-accepted changes to the original methods, including changes recommended from this review, and specific advice for tests with diluted bitumen products.

5.1 COMBINING STANDARD TESTS WITH TESTS THAT SIMULATE SITE-SPECIFIC CONDITIONS

As discussed in Chapter 3, toxicity tests fall into two major categories according to their primary objectives. Some studies compare toxicity among different types of oil, test species, and environmental conditions (e.g., temperature, salinity), and rely on standard methods to enable such comparisons. Such tests are common in regulatory testing. For example, a standard test may include a range of test concentrations in which the highest concentration far exceeds any oil concentration that might be measured at a spill. The purpose is to induce a wide range of responses so that a reliable exposure-response curve and end-point such as an LC50 or EC50 can be calculated. If only 'realistic' concentrations are tested, responses may not exceed 50%, precluding the statistical derivation of an end-point. Although oil toxicity tests conducted under 'realistic' conditions may support site-specific risk assessments, data unique to each site may not be comparable among labs using alternative test conditions. When each test is 'unique', as illustrated by the diversity of test methods catalogued in Chapter 4, the results cannot be applied broadly to create a more general understanding of hazard and risks.

The utility of ‘realistic’ tests can also be challenged because it is impossible for a lab to test every combination of conditions encountered at different spill sites. For example, CROSERF aimed to develop standardized test methods while at the same time providing test scenarios that were more realistic models of a marine spill. CROSERF methods have been described as useful in comparing oil toxicities but limited in their representation of real world environmental scenarios (Bejarano et al. 2014; Coelho et al. 2013). This is particularly the case given the new reality of constant and prolonged discharges of oil from deep sea well blow-outs (2010 Deepwater Horizon Spill), including the prolonged application of dispersants and the widespread distribution of deep water plumes of dissolved and particulate oil (Beyer et al. 2016). This scenario is of obvious concern in Canada due to offshore oil and gas development off the East coast. Other emerging scenarios that are not well understood include spills of oil and dilbit to freshwater and Arctic ecosystems.

Box 7. Combining standard and site-specific tests

For any new oil, test species or site-specific condition:

Apply standard tests to generate reference values for acute and chronic toxicity

- Follow a standard test protocol with sufficient flexibility for:
    - Marine or fresh dilution water
    - Widely-used marine or freshwater test species
    - Test conditions optimized for widely-used test species
- Estimate toxicity from measured concentrations of oil in test solutions
- Derive estimates of toxicity at site-specific exposure times by frequent measures of responses within a standard exposure duration (Sprague 1969)
- Transfer surviving organisms to freshwater to observe latent or delayed effects.
- Generate perspective by comparing toxicity to literature values for other oils tested under standard conditions, and to species sensitivity distributions

Run parallel site-specific or ‘realistic’ tests

- Apply standard methods, but systematically assess site-specific toxicity by changing one or more test conditions (e.g., oil weathering, salinity, temperature, UV light intensity, species)
- Estimate toxicity from measured concentrations of oil in test solutions
- Generate perspective on the importance of site-specific conditions by comparing toxicity to the results of parallel standard tests

One strategy to resolve the conflicting aims of tests under standard conditions and tests under site-specific conditions is to do both (Box 7). Systematic comparisons of toxicity under standard and non-standard conditions can provide data on the toxicity of a specific oil and on the relative importance of site-specific conditions in determining the relative risk of toxicity. As Box 7 indicates, site-specific data, such as toxicity over different exposure times, can also be derived from standard tests. While this approach would be more costly than single tests, it would greatly increase the ability to apply standard test data to site-specific ERAs and EIAs and to generate a more thorough understanding of which site-specific factors have the greatest effect on toxicity.

5.2 PRACTICALITY OF TEST METHODS

Not all species, exposure regimes/durations, and environmental conditions can be assessed and not all test methods and measurements are practical. Careful consideration of the time, personnel, and cost for parameters recommended is required in the development of test protocols. There is a trade-off between the amount of information required by regulators and the quality of research that can be conducted within time and budget constraints. Therefore protocol recommendations should be flexible to allow for tests of different environmental conditions typical of site-specific spills.

This review found a diversity of methods used to measure the toxicity of oil to fish, including a wide array of interpretations of previous methods. For example, numerous papers cited Anderson et al. (1974) and the CROSERF methodology (Aurand & Coelho 2005) as the source of their methods, but many of these same publications adapted the methods without providing the details or rationales for adaptations. Several reviews suggested changes to the CROSERF method (e.g., Barron & Ka'aihue 2003; NRC 2005) that have subsequently been adopted in many studies, for example to improve data for ERAs in non-temperate environments (e.g., Gardiner et al. 2013). The diversity of methods was also related to research versus regulatory testing. In reviewing the literature, George-Ares and Clark (2000) attributed variability in dispersant toxicity to the reality that not all studies were conducted to support decision making for dispersant use.

It is often difficult to compare the results of tests among studies, particularly when standard methods are not followed, or where there is little documentation of methods that could potentially change the outcome of toxicity tests. A major theme in this review is the significant difference in acute and chronic toxicity between weathered and unweathered oils, and between undispersed and chemically-dispersed oils when toxicity is expressed as the dilution of WAF or CEWAF of oil prepared with different methods. Without standard methods, and without detailed documentation of methods, it is often difficult to understand the reason for the nature and size of these differences. There is an emerging consensus that many of the differences among toxicity test data can be attributed to differences in concentrations of specific components of oil associated with acute and chronic toxicity. In cases where toxicity is expressed on the basis of measured concentrations of hydrocarbons, solutions prepared with diverse methods have equivalent toxicities. Accordingly, following a standard method to the letter is not essential if the method is described in detail and if it includes measurements of oil in water. Similarly, exposure-response relationships based on measured concentrations provide a more reliable basis for interpreting monitoring data from oil spills, or projected environmental concentrations from oil-fate models.

While environmental analytical chemistry can improve understanding of toxicity and environmental impacts, it is a limiting factor in most toxicity tests. The numbers of analyses required to characterize steady-state or time-varying concentrations of oil in water and the costs of analyses by GC-MS or GC-FID are often prohibitive. Developing alternative analytical methods for oil (e.g., fluorescence-based methods) that are cost-effective and that generate results that are closely correlated to results of conventional analyses is a critical research need. Within a study, however, it is essential to include some GC-MS or GC-FID analyses to characterize the relationship between the alternative method and conventional methods for the oil components of interest. There is also a need to develop a suite of well-characterized positive control substances (e.g., petroleum hydrocarbons known to cause acute lethality) to validate acute toxicity tests with oil.

Priority Research Need

- Develop a program to identify and disseminate reference standards and reference oils.

5.3 PROPOSED FRAMEWORK

In response to these issues, a proposed framework of recommended test methods was compiled from Appendix 1 that summarizes the CROSERF protocols for marine oil toxicity tests with WAF and CEWAF (Table 7) and suggests amendments to respond to testing issues identified in Chapters 2 -4. The summary makes a distinction between acute and chronic toxicity tests, and is divided into sections corresponding to the requirements for:

- Explicit statements of the objectives of the test
- Experimental design
- Toxicity testing
- Physically and chemically characterizing the test oil
- Chemically characterizing test solutions
- Measuring responses and calculating endpoints; and
- Guidance on reporting the results.

For each method, criteria are presented that should be considered when testing any oil product. Nevertheless, test methods should be sufficiently flexible to allow their application to site-specific conditions or to oils with unique characteristics, particularly if tests are backed-up by sufficient chemical analyses to understand the relationships between exposure and effect common to all oils. Many of the recommendations for test methods in Table 7 are unchanged from the original CROSERF guidelines (Aurand & Coelho 2005). Where changes have been made, or there is a change in emphasis, they are highlighted in bold, and the rationale for these changes is explained in the methods reviews of Chapter 4. These modifications also apply to other petroleum products with extreme or unique characteristics, such as HFO or Bakken light crude.

As mentioned in Section 3.1, the requirements of oil toxicity tests may also conflict with established methods. For example, one area identified in Chapter 4 where test methods should be modified from standard guidance is a recommendation for all-glass or stainless steel containers to avoid significant partitioning of hydrocarbons from test solutions to the plastic chambers currently recommended for early life stage toxicity tests (EC 1998). Similarly, change is needed to Environment and Climate Change Canada guidance that chemical analyses are optional when there are limitations in analytical techniques, high costs, or previous information is available (EC 1998; 2007; 2011). Unless oil toxicity is expressed on a common basis of measured concentrations of hydrocarbons, the results of toxicity tests will be misleading, representing a waste of testing resources, and potentially very expensive errors if the data are used for ERAs or EIAs. Section 4.4 provides cost-effective tools and strategies for chemical characterization of test solutions.

Table 7: A proposed framework for toxicity tests of physically- and chemically -dispersed oil to fish to support regulatory assessments and comparisons of toxicity among oils, species, and test conditions. WAF refers to the water accommodated fraction of oil produced by low, medium, or high energy mixing of oil and water, modified from the CROSERF methods. Similarly, CEWAF refers to chemically-enhanced WAF produced by low, medium or high energy mixing. Where changes have been made to the CROSERF methods, or there is a change in emphasis, they are highlighted in bold, and the rationale for these changes is explained in the methods reviews of Chapter 4. If the properties of test oils (e.g., viscosity) might interact with test methods (e.g., oil dispersion) and affect the efficiency of mixing or the outcome of toxicity tests, preliminary experiments are recommended to indicate whether changes to methods are needed to avoid bias. It is also recommended that these changes be described in detail and justified.

1. **Experimental Design**

<table>
<tr><th rowspan="2"></th><th colspan="2">Nature of test solution</th></tr>
<tr><th>Physically- dispersed (WAF)</th><th>Chemically- dispersed (marine, CEWAF)</th></tr>
<tr><td>1.1. Test species, life stage</td><td colspan="2">• **If possible, ECCC, EPA, OECD standard test species and life stages or commonly used ‘model’ species**
• **If not, justify**</td></tr>
<tr><td>1.2. Concentration range</td><td colspan="2">• Encompass range from non-toxic to toxic to ensure that end-points can be calculated</td></tr>
<tr><td>1.3. Exposure time</td><td colspan="2">• Acute – **Follow standard test protocol (See 1.1), typically 96 h**
• Chronic – **Follow standard test protocol (See 1.1)**
• **If end-points are needed for shorter exposure times (e.g., 12 h LC50s), make observations of responses at short intervals (e.g., 1,2,4,8,16 h from start)**</td></tr>
<tr><td>1.4. Exposure regime options</td><td colspan="2">• Acceptable test regimes include: static non-renewal (SNR), static renewal (SR), and continuous-flow to recognize the objectives of the test.</td></tr>
<tr><td>1.5. Water type/quality</td><td colspan="2">• **Fresh water to** marine, characterized by chemical analysis
• Fresh water: characterize **pH, alkalinity, conductivity, hardness, TOC**
• Salt water and brackish water: characterize **salinity or conductivity, pH, TOC**</td></tr>
<tr><td>1.6. Oil state</td><td>• Weathered or unweathered oil</td><td>• Dispersed weathered or unweathered oil</td></tr>
</table>

1.7. Controls	• Water only • **Positive controls are recommended, where appropriate** • **For example when testing a previously untested oil include a well-tested oil as a reference toxicant (e.g., ANSC, PBCO, MESA)**	• Water only • **Positive controls are recommended, where appropriate** • **For example when testing a previously untested oil, include a well-tested oil as a reference toxicant (e.g., ANSC, PBCO, MESA)** • **Dispersant control (e.g., Nujol CEWAF at same OWR and DOR as oil test)**

2. Preparing Test solutions

	Nature of test solution	
	Physically- dispersed (WAF)	**Chemically- dispersed (marine, CEWAF)**
2.1. Oil and dispersant storage and handling	• Fresh stocks of oil should be thoroughly mixed in a closed container and stored in tightly sealed glass or metal containers with minimal headspace. **Refer to ASTM D4057 for volatile petroleum handling recommendations.** • **Oil should be characterized physically and chemically when received, when first used if stored longer than 1 month, and at the end of experiments if they last for more than 1 month** • **Aliquot fresh oil into volumes needed for each days' use, and store in sealed containers with minimum head-space at 4ºC; re-mix if stored longer than 1 month** • **Store as briefly as possible before testing and open bottles as infrequently as possible** • **The same procedures should be used for fresh stocks of dispersant, but without physical or chemical characterization**	
2.2. Oil weathering	• **ASTM Method D2892 or D86 methods can be used to artificially weather oils, but temperatures above 130˚C must not be used to avoid chemically changing the oil (including dehydration of cycloalkanes and breaking of chemical bonds). Under vacuum distillation conditions (ASTM Method D1160), maximum recommended temperature is 200˚C.** • **Weather to a constant weight or target loss, not to exceed 48 h**	

	Nature of test solution	
	Physically- dispersed (WAF)	**Chemically- dispersed (marine, CEWAF)**
2.3. Dilution water	• Use the same water source for holding and acclimating stocks of fish, for preparing WAF/CEWAF, and for preparing test solutions • **Fresh water: characterize pH, alkalinity, conductivity, hardness, TOC** • Salt water and brackish water: characterize **salinity or conductivity, pH, TOC** Marine tests: • Minimum filtration 0.45 μm • Local sea water is recommended; dilute with deionized water as needed to achieve a target salinity • Reconstituted sea water is acceptable if it supports good survival and health of test organisms Fresh water tests: • **Minimum filtration 0.45 μm if derived from an untreated surface water** • **For municipal water sources, treat to reduce total chlorine to <10 μg/L**	
2.4. Dissolving oil in water	• Batch: Stirring, mixing • **Continuous-flow: pumps, oil desorption columns**	• Batch: Stirring, mixing with chemical dispersants • **Continuous-flow: pumps**
2.5. Oil-water ratios (OWR)	• **Acute and chronic toxicity tests – 25 g/L or sufficient oil that concentrations of oil in WAF and CEWAF are not limited by the amount of oil added, as indicated by residual floating oil after mixing, and by chemical analysis of stock solutions.**	
2.6. Mixing temperature	• **Mix at room temperature (20-22˚C)**	
2.7. Mixing light	• Mixing should be carried out in darkness **or in low UV laboratory fluorescent lighting**	
2.8. Addition of oil	• **Add known quantity of oil at centre of the water surface during mixing** • Calculate delivery mass by weight difference	
2.9. Mixing vessel size	• **No restrictions**	
2.10. Headspace	• **Acute toxicity tests – 20-25% headspace in a sealed mixing vessel** • **Chronic toxicity tests – open mixing vessel**	

	Nature of test solution	
	Physically- dispersed (WAF)	**Chemically- dispersed (marine, CEWAF)**
2.11. Mixing energy	- When using a magnetic stir bar the amount of energy applied is judged solely by the depth of the mixing vortex - **For both WAF and CEWAF, adjust stirring to give a mixing vortex equivalent to 20-25% of water depth (Clark et al. 2001)** - **Mixing methods with higher energy and chemical dispersion may be beneficial when mixing oils with higher viscosity** - **For HEWAF and HECEWAF, use vortex mixer and sonication to make a thorough mix of oil and water** - Conduct a pre-test on oil to determine how the oil and dispersant should be added to the water, and whether oil-in-water emulsions form. If the latter, reduce the mixing energy or try a different OWR - **When available, use standard protocol/apparatus with known energy dissipation rate**	
2.12. Mixing duration	- Should be **18 h to suit the 24 h cycle of daily solution preparation and to avoid bias by microbial growth**	- Total stirring time should be 18 h to **suit the 24 h cycle of daily solution preparation** and to avoid bias by microbial growth. **Add dispersant immediately after oil vortex is established**.
2.13. Dispersant- oil ratio (DOR)	- N/A	- **1:20**
2.14. Addition of dispersant	- N/A	- Add known quantity of dispersant to oil at the centre of vortex after a vortex has been established - Calculate delivery mass by weight difference - Do not pre-mix oil and dispersant
2.15. Settling times	- **Settling for 6 h prior to extraction of aqueous layer to allow resurfacing of oil droplets** - **Use immediately – do not store prior to use**	
2.16. Dilution gradients	- Prepare gradients with variable dilution of WAF or CEWAF stock solutions, **or by preparing a series of oil desorption columns with gradients of oil concentrations in gravel, or by variable dilutions of a continuous flow of oil-contaminated water produced by oil-pumping and mixing systems** - Serial dilutions of a stock solution are not recommended because of a potential bias in the relative proportions of hydrocarbons with different water solubilities	
2.17. Storage of test solutions	- **Do not store test solutions**	

3. Toxicity Testing

	Nature of test solution	
	Physically- dispersed (WAF)	**Chemically- dispersed (marine, CEWAF)**
3.1. Test species	• Guidelines for fish culture and handling should follow those published by one of ECCC, ASTM, OECD, EPA, etc., and local animal care protocols • **When fish are fed, daily solution renewals should immediately follow feeding, preferably by transferring the fish to a fresh test solution in a clean tank. The previous tank should be thoroughly cleaned to remove debris and organic films that absorb hydrocarbons or harbour bacteria that can create metabolites of hydrocarbons.** • **Remove dead animals daily to avoid accumulation of biological oxygen demand (bacteria) and organic matter that could absorb hydrocarbons**	
3.2. Biomass loading	• **Follow the standard protocol for each test organism**	
3.3. Test solution volume and container size	• **Follow the standard protocol for each test organism**	
3.4. Test container material	• **All glass (seamless, i.e., no aquaria with plate glass assembled with silicone seal) or stainless steel** • **Avoid plastics, with the exception of Teflon**	
3.5. Aeration	• **Acute - Aeration of acute lethality tests should be minimized to reduce volatilization and weathering of compounds causing acute narcosis, but sufficient to sustain > 60% saturation (EPA 850.1075)** • **Chronic - Low rates of aeration sufficient to maintain oxygen concentrations within the optimal range for each test species**	
3.6. Test containers	• Acute - Closed container to retain volatiles (covered but not air-tight seal) • **Chronic - Open container**	
3.7. Endpoints	• Lethality (i.e., 96 h LC50) • “Initial effect” narcosis/moribundity (0.5 to 1 h EC50) • **Sublethal toxicity (EC50; e.g., embryo toxicity, growth, survival, behaviour, reproduction, molecular, physiological responses)**	

4. Characterizing oil

	Nature of test solution	
	Physically- dispersed (WAF)	**Chemically- dispersed (marine, CEWAF)**
4.1. Chemical composition	• **Saturates, aromatics, resins, and asphaltenes (SARA)** • **VOCs including BTEX, TPH (CCME fractions) (CCME 2001)** • **PAH and TPAH (See minimum analyte list in Appendix 2, Table A2.2)** • **If possible, characterize oil hydrocarbons by high-temperature simulated distillation (e.g. ASTM D7169)**	
4.2. Physical characteristics	• **Viscosity, density**	
4.3. Storage and Handling	• Store in tightly sealed glass or metal containers with minimal headspace • **Open containers as infrequently as possible (See section 2.1)** • Store in darkness at **4˚C** • Recommended chemical characterization of stored oil to monitor changes in composition over time	

5. Characterizing test solutions

	Nature of test solution	
	Physically- dispersed (WAF)	**Chemically- dispersed (marine, CEWAF)**
5.1. Sampling, Storage and Handling	• **Water samples for hydrocarbon analyses must be collected from fresh WAF and CEWAF stock solutions prior to toxicity testing** • **Water samples for hydrocarbon analyses should be collected from each test solution at the start and end of a toxicity test (preferably more frequently), to compare measured concentrations of oil to nominal dilutions.** • **The highest concentrations of test solutions should be sampled frequently (N≥3 at reasonable intervals throughout test) to describe temporal trends of oil concentrations and to calculate mean concentrations with a measure of variance.** • **Pooling should be considered where test solution volumes are restricted** • **For static, or static daily renewal exposures, the highest concentration should be sampled repeatedly over a 24-hour cycle of solution renewals to describe the time-varying exposure regime.** • **If samples are analyzed externally by a qualified oil chemistry lab (recommended), consult with the lab in advance about sampling, pooling sample volume, bottles, preservatives, storage conditions, storage time, shipping, and selection of internal standards and recovery standards.** • **Samples should be processed (e.g. extracted) at the laboratory within 14 days of collection, and analyzed within 14 days of processing. Samples should be held at 4 ^{0}C during shipping and storage.**	
5.2. TPH ($C_{10} – C_{36}$)	• **Analyze TPH from CCME fractions (CCME 2001)** • Do not correct for analytical recovery • Detection limit of at least 10 ppb for any specific n-alkane that is in the standard	
5.3. VOC including BTEX ($C_6 < C_{10}$)	• **See Appendix 2 Table A2.2 for list of minimum target analytes** • **VOCs should be measured following the USEPA method EPA SW-846**	
5.4. Semi-volatiles and PAHs	• **See Appendix 2, Table A2.2 for a minimum list of PAH, including alkylated homologs** • **TPAH as defined as the sum of PAH from the recommended list** • Do not correct for analytical recovery • Detection limit of at least 10 ppb for any specific PAH that is in the standard	

	Nature of test solution	
	Physically- dispersed (WAF)	**Chemically- dispersed (marine, CEWAF)**
5.5. Rapid oil analysis (e.g., fluorescence and UV detection)	• **Methods for semi-quantitative rapid analyses can supplement other quantitative analytical methods for repeated characterizations of WAF, CEWAF and test solutions provided that a sub-set of parallel samples are analyzed by the quantitative methods.** • **Perform calibration with standard solutions as available, using the same matrix as the test solutions.**	

6. Measuring Responses, calculating end-points

	Nature of test solution	
	Physically- dispersed (WAF)	**Chemically- dispersed (marine, CEWAF)**
6.1. Statistics	• **Must include statistical analysis of data, including variability (e.g., LC50s and EC50s including confidence limits by statistical tests appropriate for binomial or continuous data)** • **Express endpoints in terms of measured TPH and TPAH, or other analytes as appropriate**	

7. Guidelines for Reporting; where methods deviate from the standard, these changes should be reported (See Appendix 4 for additional guidance)

	Nature of test solution	
	Physically- dispersed (WAF)	**Chemically- dispersed (marine, CEWAF)**
7.1. Objectives of the test	• **Explicitly state the objectives of the test**	
7.2. Rationale and details for the experimental design	• **Explain how the experimental design addresses the test objectives** • **Identify all controls and a rationale for including each** • **Provide details on the ranges of test concentrations, the intended pattern of exposure (e.g., constant or declining concentrations), responses measured, and sample sizes** • **Identify assumptions inherent in methods applied for testing toxicity and characterizing test solutions, and discuss the importance of any violations of these assumptions**	

	Nature of test solution	
	Physically- dispersed (WAF)	Chemically- dispersed (marine, CEWAF)
7.3. Rationale and details for methods	• **Outline all methods for preparing test solutions, toxicity testing, chemical analyses of oil and test solutions, data manipulation, and statistical analyses** • **Provide a citation to any published protocols that were followed, identify any deviations from the published protocol, and provide a rationale for any changes** • **Provide sufficient details of the test methods that the experiment can be replicated by another laboratory**	
7.4. Characteristics of the test oil	• **Results of physical and chemical characterization of the oil when received, when tested, and when testing was complete, with an emphasis on any loss of volatiles due to weathering, concentrations of low molecular weight compounds (<500) for acute lethality tests, and TPH and TPAH concentrations for chronic toxicity tests** • **For artificially weathered oils, the maximum temperature of the method used should be reported**	
7.5. Characteristics of test solutions	• **Report the sampling design** • **Compare measured concentrations (with sample size and a measure of variance) to nominal concentrations of oil** • **Report other water quality characteristics, including average (variance, *N*) of salinity, alkalinity, temperature, pH, oxygen**	
7.6. Variability in test solution concentration	• **Measured oil concentrations over the duration of the experiment, including changes in concentrations over the intervals between solution renewals for SR protocols. Graph the decay rate of measured concentrations in test solutions over 24 h, and over the course of the entire experiment, with a measure of variance and *N*.** • **For acute lethality tests at T=0 and at the end of the test, report measures of volatiles and low molecular weight compounds (<500) to demonstrate the extent of losses due to weathering** • **For chronic toxicity tests, report the average (variance, *N*) of measured concentrations of oil over time of the highest, mid-level and control concentrations** • **Compare measured concentrations to nominal dilutions**	
7.7. Characteristics of test organisms	• **Source, age, and life stage of test organisms and whether they were from an aquaculture facility or an in-house culture** • **Range and average weight of test organisms and biomass loading rate in test solutions (i.e., g biomass/L of test solution)** • **Mortality rates of the stock population in the week prior to testing** • **Mortality rates in each oil concentration, including controls, throughout the test** • **Any unusual behaviour or evidence of stress or poor health**	

	Nature of test solution	
	Physically- dispersed (WAF)	Chemically- dispersed (marine, CEWAF)
7.8. Statistics	• **Report statistical methods for estimating LC50, EC50, or other toxicity data** • **Show all results using measured concentrations of hydrocarbons as the metric of exposure** • **Include sample size and a measure of variance (preferably 95% confidence limits)** • **Note and justify any data manipulations, including identification and removal of outliers and calculation of ratios, and note any potential for statistical bias.**	

While the use of standard methods and protocols is highly recommended for toxicity tests that will contribute to ERAs and EIAs, as outlined in Table 7, it is recognized that some tests will have site-specific objectives that require deviations from the range of parameters suggested by the standard framework. When there are deviations from the standard method, detailed reporting of the parameters of the test are required and it is recommended that the tests be run in parallel with the standard method (See Section 5.1). Test parameters where there may be deviations from the standard method for site-specific or species-specific assessments may include the test species, life stage, water type/quality, oil state, and/or temperature.

5.4 NEXT STEPS: DEVELOPING A STANDARD METHOD BY COLLABORATION AMONG SECTORS

To translate the proposed framework for oil toxicity testing into one or more standard methods, a consensus-based process involving scientists from industry, government and academia is recommended, similar to the CROSERF initiative. Such a development would complement past initiatives of Environment and Climate Change Canada's (ECCC) Biological Methods Development Group. One of the key aspects of the CROSERF initiative was the ability to bring together industry, academia, and government to facilitate open discussion and form collaborations. The discussion of different parties' concerns about test methods improved the acceptance of the standards generated, while broadening the applicability of the methods developed. Additionally, collaboration reduced the cost and the number of studies for external confirmation while facilitating comparability among studies. With widely accepted methods, more time is spent discussing the implications of measured toxicity than how the test methods affected measured toxicity. The end goal is to develop a standard method to assess oil toxicity, where the data generated is comparable to support decision-making

6 RESEARCH RECOMMENDATIONS

The extensive review of the effects of oil spills by the Royal Society of Canada (Lee et al. 2015) identified research needs related to understanding the fate, behaviour and effects of oil in Canada's aquatic ecosystems, particularly in fresh and Arctic waters. The RSC review identified research needs for dilbit as a unique oil product, as did the National Academy of Sciences (NAS 2016). Dew et al. (2015) identified knowledge gaps for dilbit spills in fresh water relating to its fate, the ranking of species sensitivity to toxicity, the effect of environmental factors on toxicity, and the mechanisms of toxicity. In this review, research needs are restricted to methods for oil toxicity testing and assessment of factors affecting the outcomes of toxicity tests, particularly with dilbit. These needs have been consolidated from the text of this review and from the reviews cited above. Some of the identified research needs below are currently the subject of ongoing research.

Priority Research Needs List:

- Assess the effect of all aspects of solution preparation (e.g., mixing vessel headspace; OWR, test volume) on the chemical composition, concentrations and toxicity of WAF and CEWAF.
- Developing a routine and practical method (or system) for determining mixing energy during solution preparation by stirring, sonication, high energy mixing, etc.
- Conduct a literature survey to demonstrate the range of mixing energies associated with spills to open oceans, coastlines, lakes, and rivers of different current strengths and turbulence.
- Develop a program to identify and disseminate reference standards and reference oils.

- Develop rapid and inexpensive analytical methods to chemically characterize oil and test solutions including SARA analysis and high temperature distillation techniques.

Preparation of test solutions

- Assess the interactions of dissolved and particulate oil (physically- and chemically-dispersed) with dissolved and particulate organic matter, and their effects on the measured toxicity of oil.
- Develop inexpensive and practical methods for measuring the concentration and size distribution of oil droplets in test solutions to distinguish dissolved from particulate oil.
- Examine the effect of mixing temperature on the chemical composition, concentration, and toxicity of WAF and CEWAF.
- Identify experimental factors that affect the composition and toxicity of test solutions prepared with oil-desorption columns and high pressure pumps and mixers.
- Assess the duration of mixing needed for dilbits with different formulations and weathering states to reach functional equilibria between water and oil phases, particularly across a range of temperatures.

Performance and related experimental details

- Develop data quality objectives to assess the reliability of chemical analyses of oil and of test solutions and to compare results and performance among multiple laboratories.
- Conduct inter-laboratory toxicity testing of shared samples of dilbit and crude oil to estimate the variance in test results that should be incorporated as uncertainty in ERAs and EIAs, and to identify the test variables that contribute most strongly to variance among labs.
- For the proposed test framework, test the assumptions inherent in aquatic toxicity tests of oil to determine which are met and which are violated.

Exposure details

- Optimize the application of passive samplers for characterizing concentrations of dissolved hydrocarbons in test solutions.
- Assess the extent to which dilbit products are photosensitized in laboratory tests and under the range of UV intensities typical of Canadian environments.
- Develop practical methods for measuring “adhesion” of oils.
- Develop statistical models that integrate oil concentrations in test solutions over time to generate a stable metric of exposure and calculated toxicity.
- Identify and evaluate practical and cost-effective test protocols to maintain constant concentrations of oil in test solutions over the duration of toxicity tests.
- Identify the most suitable toxicity test vessels, assess the absorption of oil from test solutions to various tank materials and the extent, time-course, and nature of changes in the chemical composition of test solutions.

Dilbit and other oils

- Characterize the nature and extent of physical and chemical changes to test oils under different storage conditions, and the potential effects on the outcome of toxicity tests.
- Compare the behaviour and toxicity of dilbit among marine, brackish and fresh waters, including those that are rich in humic acids or very low in alkalinity (also hardness, pH, other parameters), and under temperate and Arctic conditions.

- Determine the effects on oil toxicity of variations in key water quality parameters (e.g., pH, alkalinity, salinity, DOC, POC, humic acids).
- Compare the toxicity of dilbit to that of previously-tested conventional crude oils, using the same standard methods for preparing test solutions and measuring toxicity.
- Develop standard methods to assess the toxicity of dilbit incorporated into sediments.
- Apply effects driven chemical fractionation (EDCF) to identify the toxic components of oil.
- Develop species sensitivity distributions based on standard oil toxicity tests.
- Assess the contributions to toxicity of products of photodegradation and biodegradation of hydrocarbons, including acid-extractable organics.
- Repeat toxicity tests with various dilbit products and test species typical of Canadian environments, using experimental designs that adhere to standardized methods.

Oil Weathering

- Test light and heavy crude oils at multiple weathering states to determine how chemical composition and toxicity of test solutions changes with mixing times and temperatures at different times during testing.
- Assess the effects of oil weathering during solution preparation on the density (API Gravity) and viscosity of test oils, on the chemical composition of test solutions prepared by standard methods, and on the measured toxicity of oil.
- Assess the effects of sample preparation on dilbit weathering during toxicity testing, and develop procedures for dilbit sample preparation that minimize weathering.

Spill treating agents

- Identify the optimal dispersant-oil ratios and methods for adding dispersant to oil to minimize ‘free’ dispersant in test solutions, to maximize the amount of dissolved and particulate oil in solution, and to minimize dispersant toxicity in dispersed oil toxicity tests. For risk assessments, tests of different ratios may be needed to match site-specific conditions or applicable regulations.
- Assess different controls for tests of chemically-dispersed oil to identify whether or not dispersant affects the outcome of toxicity tests, and whether there is oil-dispersant interactive toxicity.
- Assess the effectiveness of chemical dispersion of weathered and unweathered dilbit according to the size spectra of droplets produced, the chemical composition of test solutions, and measured toxicity.
- Assess the chronic and sublethal effects on fish of chemical dispersants, both alone and in combination with oil.
- Develop methods to characterize ‘free’ dispersant in test solutions and dispersant associated with oil particles.

7 CONCLUSIONS

This review of the literature on methods to measure the toxicity of oil to fish demonstrated the need to develop more standardized test methods to support ERAs and EIAs. The literature is characterized by a high diversity of test methods, oils tested, and test conditions, most of which have not been examined systematically to demonstrate how they affect estimated toxicity. Too many reports provide too little detail on the oils tested and the test conditions to fully interpret the test data. Thus, comparisons of toxicity among oils, among test species, and among test conditions are confounded and the data are often inappropriate for site-specific ERAs and EIAs.

These issues are particularly important for dilbit products which have unique properties compared with conventional crude oils including rapid weathering, a predominance of higher molecular weight hydrocarbons such as PAH, and greater adhesion. These unique properties have a strong influence on the behaviour and chemistry of dilbit during testing and on its perceived toxicity. The same would also be true of the ultra light oils typical of fracking, and the heavy oils that will predominate as conventional crude supplies decline. *Based on this review, experiments testing new or unusual petroleum products should conduct preliminary tests to determine how the properties of these oils influence their mixing characteristics and the transfer of oil to water. Based on these experiments, small changes to standard protocols may resolve any issues and should be reported in detail.*

Overall, this review may give the false impression that all toxicity test data are unreliable. However, the opposite is true: a large portion of the studies reviewed represented good science, but the site-specific conditions of these tests limited their applicability to EIAs and ERAs of oil spills under different circumstances. The data may be useful, but expert judgment is needed to understand how reported toxicity might change with circumstances. It must also be recognized that toxicity tests in which oil concentrations decline with time, and chemical analyses of test solutions that do not discriminate between dissolved and particulate hydrocarbons, likely underestimate the true toxicity of oil dissolved in water and contribute to the uncertainty of estimated risks.

An attempt to alleviate these issues with standardized test methods (CROSERF) specified: analyses to characterize test oils and test solutions; protocols for preparing solutions of WAF and CEWAF, for exposing aquatic organisms to oil solutions over 96 hours, and for measuring responses; and reporting requirements to detail test conditions and toxicity data. However, the basis of the CROSERF method was a marine oil spill scenario typical of a ship sinking in the open sea, with rapid dissipation of oil concentrations in surface waters and measurement of acute (<96) lethality as concentrations declined. Scenarios of constant discharges to deep water over prolonged periods (e.g., Gulf of Mexico Deepwater Horizon spill, 2010), spills to confined freshwater lakes and rivers (e.g., Wabamun Lake spill, 2005; Kalamazoo River spill, 2010), and chronic exposure and toxicity were not considered.

The alternative to the site-specific approach is to develop more generic methods that generate toxicity data reflecting primarily the chemical composition of the test oil and not its physical characteristics or unique test conditions that modify the extent to which aquatic organisms are exposed to oil. A standard method should generate toxicity data that can be compared among oils of widely different types and a relative hazard ranking among oils based on their chemistry. The application of standard data for ERAs and EIAs of site-specific spill conditions would use models of oil fate and behaviour to predict the extent of exposure of aquatic species to specific constituents of oil, and hence toxicity. More detailed information to validate model predictions could be derived from studies comparing the effect of different environmental conditions on fate, behaviour and toxicity of common oils, or by comparing tests under site-specific conditions coupled with tests under standard conditions to provide a reference.

This review presents a proposed framework for standardized tests that lists the detailed requirements of the CROSERF method, amended with recommendations for improved definitions of test objectives, experimental designs, methods for preparing test solutions of physically- and chemically-dispersed oil, chemical analyses to characterize test solutions, exposure regimes for acute and chronic toxicity tests, and statistical methods and reporting requirements for test conditions and results. This is not a standard method, but rather a flexible framework to guide future oil toxicity tests, to identify research needs for improving test methods, and to stimulate method development for different aspects of standard toxicity tests. One of the most valuable aspects of the CROSERF process was the collaboration of government, university and industry scientists to develop consensus-based recommendations for test methods, and this approach is recommended to build on the results of this review.

8 LITERATURE CITED

Adams, J., Bornstein, J.M., Munno, K., Hollebone, B., King, R.T., Brown, S., and Hodson, P.V. 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., Sweezey, M., and Hodson, P.V. 2014b. Oil and oil dispersant do not cause synergistic toxicity to fish embryos. Environ. Toxicol. Chem. 33: 107–114.

Adeyemo, O.K., Kroll, K.J., and Denslow, N.D. 2015. Developmental abnormalities and differential expression of genes induced in oil and dispersant exposed *Menidia beryllina* embryos. Aqua. Toxicol. 168: 60–71.

Agamy, E. 2013. Impact of laboratory exposure to light Arabian crude oil, dispersed oil and dispersant on the gills of the juvenile brown spotted grouper (*Epinephelus chlorostigma*): A histopathological study. Mar. Environ. Res. 86: 46–55.

Akaishi, F.M., Silva de Assis, H.C., Jakobi, S.C.G., Eiras-Stofella, D.R., St-Jean, S.D., Courtenay, S.C., Lima, E.F., Wagener, A.L.R., Scofield, A.L., and Oliveira Ribeiro, C.A. 2004. Morphological and neurotoxicological findings in tropical freshwater fish (*Astyanax* sp.) after waterborne and acute exposure to water soluble fraction (WSF) of crude oil. Arch. Environ. Con. Tox. 46: 244–253.

Alderman, S.L., Lin, F., Farrell, A.P., Kennedy, C.J., and Gillis, T.E. 2016. Effects of diluted bitumen exposure on juvenile sockeye salmon: From cells to performance. Environ. Toxicol. Chem. doi: 10.1002/etc.3533.

Alloy, M., Baxter, D., Stieglitz, J., Mager, E., Hoenig, R., Benetti, D., Grosell, M., Oris, J., and Roberts, A. 2016. Ultraviolet Radiation Enhances the Toxicity of Deepwater Horizon Oil to Mahi-mahi (*Coryphaena hippurus*) Embryos. Environ. Sci. Technol. 50(4): 2011–2017.

Anderson, J.W., Neff, J.M., Cox, B.A., Tatem, H.E., and Hightower, G.M. 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. 2014. Standard Guide for the Use of Lighting in Laboratory Testing, E1733-95, p. 1279–1289, in: 1996 Annual Book of ASTM Standards, Vol. 11.05, *In* American Society for Testing and Materials, Philadelphia, PA (1996).

Aurand, D. and Coelho, G. 2005. Cooperative aquatic toxicity testing of dispersed oil and the Chemical Response to Oil Spills: Ecological Effects Research Forum (CROSERF). Ecosystem Management & Associates, Inc. Technical Report. 105 p.

Austad, T., Shariatpanahi, S.F., Strand, S., Black, C.J.J., and Webb, K.J. 2011. Conditions for a low-salinity enhanced oil recovery (EOR) effect in carbonate oil reservoirs. Energ. Fuel. 26: 569–575.

Bakker, G. 2011. The corrosive nature of diluted bitumen and crude oil literature review. M.R. Gordon and Associates Ltd. 16 p.

Barron, M.G. and Ka'aihue, L. 2001. Potential for photoenhanced toxicity of spilled oil in Prince William Sound and Gulf of Alaska waters. Mar. Pollut. Bull. 43: 86–92.

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

Barron, M.G., Carls, M.G., Short, J.W., and Rice, S.D. 2003. Photoenhanced toxicity of aqueous phase and chemically dispersed weathered Alaska North Slope crude oil to Pacific herring eggs and larvae. Environ. Toxicol. Chem. 22: 650–660.

Barron, M.G., Carls, M.G., Short, J.W., Rice, S.D., Heintz, R.A., Rau, M., and Di Giulio, R. 2005. Assessment of the Phototoxicity of Weathered Alaska North Slope Crude Oil to Juvenile Pink Salmon. Chemosphere. 60: 105–110.

Bauder, M.B., Palace, V.P., and Hodson, P.V. 2005. Is Oxidative Stress the Mechanism of Blue Sac Disease in Retene-Exposed Trout Larvae? Environ. Toxicol. Chem. 24: 694–702.

Bejarano, A.C., Levine, E., and Mearns, A.J. 2013. Effectiveness and potential ecological effects of offshore surface dispersant use during the Deepwater Horizon Oil Spill: A retrospective analysis of monitoring data. Environ. Monit. Assess. 185: 10281–10295.

Bejarano, A.C., Clark, J.R., and 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.

Bennett, D., Girling, A.E., and Bounds, A. 1990. Ecotoxicology of Oil Products: Preparation and Characterisation of Aqueous Test Media. Chemosphere. 21: 659–669.

Beyer, J., Trannum, H.C., Bakke, T., Hodson, P.V., and Collier, T. 2016. Environmental effects of the Deepwater Horizon oil spill: a review. Marine Poll. Bull. 110: 28–51.

Bidwell, J.R., Cherry, D.S., and Merski, A.T. 2003. Toxicity evaluation of a commercial bioremediation agent mixed with crude oil. Environ. Toxicol. Chem. 22: 84–91.

Binder, R.L. and J.J. Stegeman. 1983. Basal levels and induction of hepatic aryl hydrocarbon hydroxylase activity during the embryonic period of development in brook trout. Biochem. Pharmacol. 32: 1324–1327.

Blondina, G.J., Singer, M.M., Lee, I., Ouano, M.T., Hodgins, M., Tjeerdema, R.S., and Sowby, M.L. 1999. Influence of salinity on petroleum accommodation by dispersants. Spill. Sci. and Technol. Bull. 5: 127–134.

Bornstein, J.M., Adams, J.E., Hollebone, B., King, T., Hodson, P.V., and Brown, R.S. 2014. Effects-driven chemical fractionation of heavy fuel oil to isolate compounds toxic to trout embryos. Environ. Toxicol. Chem 33: 814–824.

Bragin, G.E., Parkerton, T.F., Redman, A.D., Letinksi, D.J., Butler, J.D., Paumen, M.L., Sutherland, C.S., Knarr, T.M., Comber, M., and den Haan, K. 2016. Chronic toxicity of selected polycyclic aromatic hydrocarbons to algae and crustaceans using passive dosing. Environ. Toxicol. Chem. 9999: 1–10.

Brooks, M. 2014. The Cautionary Tale of Kalamazoo: The still-unfolding horrors of spilled bitumen in Michigan's Kalamazoo River explain why more pipelines are an unacceptable risk. Alt. J. 40: 18–25.

Brooks, P.W., Fowler, M.G., and Macqueen, R.W. 1988. Biological marker and conventional organic geochemistry of oil sands/heavy oils, Western Canada Basin. Org. Geochem. 12: 519–538.

Camus, L., Aas, E., and Børseth, J.F. 1998. Ethoxyresorufin-O-deethylase activity and fixed wavelength fluorescence detection of PAHs metabolites in bile in turbot (*Scophthalmus maximus L.*) exposed to a dispersed topped crude oil in a continuous flow system. Mar. Environ. Res. 46(1): 29–32.

Canevari, G.P., Calcavecchio, P., Becker, K.W., Lessard, R.R., and Fiocco, R.J. 2001. Key parameters affecting the dispersion of viscous oil. *In* International Oil Spill Conference pp. 479–484.

Carls, M.G., Rice, S.D., and Hose, J.E. 1999. Sensitivity of fish embryos to weathered crude oil: Part I. low level exposure during incubation causes malformations, genetic damage, and mortality in larval Pacific herring (*Clupea pallasi*). Environ. Toxicol. Chem. 18: 481–493.

Carls, M.G. and Thedinga, J.F. 2010. Exposure of pink salmon embryos to dissolved polynuclear aromatic hydrocarbons delays development, prolonging vulnerability to mechanical damage. Mar. Environ. Res. 69: 318–325.

Carlsson, G., Norrgren, L., Hylland, K., and Tollefsen, K.E. 2014. Toxicity screening of produced water extracts in a zebrafish embryo assay. J. Tox. Env. Health A. 77: 600–615.

CBC. 2016a. Expect oil along shores of North Saskatchewan River, says spill expert. Curves of the river will send oil to shore areas, expert says. (accessed August 2, 2016).

CBC. 2016b. Natural recovery part of Husky oil spill response, expert says. (accessed August 2, 2016).

CCME. 2001. Reference Method for the Canada-Wide Standard for Petroleum Hydrocarbons in Soil – Tier 1 Method. Canadian Council of Ministers of the Environment, Winnipeg, MN. (accessed February 27, 2017).

Chandrasekar, S., Sorial, G.A., and Weaver, J.W. 2006. Dispersant effectiveness on oil spills - impact of salinity. ICES J. Mar. Sci. 63: 1418–30.

Chapman, H., Purnell, K., Law, R.J., and Kirby, M.F. 2007. The use of chemical dispersants to combat oil spills at sea: A review of practice and research needs in Europe. Mar. Poll. Bull. 54: 827–838.

Choyke, S and Ferguson, P.L. 2015. Analysis of Corexit 9500 dispersant biodegradation in seawater using high resolution mass spectrometry. SETAC, Salt Lake City.

Clark, J., Becker, K., Venosa, A., and Lewis, A. 2005. Assessing dispersant effectiveness for heavy fuel oils using small-scale laboratory tests. *In* International Oil Spill Conference Proceedings: 59–63.

Clark, J.R., Bragin, G.E., Febbo, E.J., and Letinski, D.J. 2001. Toxicity of physically and chemically dispersed oils under continuous and environmentally realistic exposure conditions : Applicability to dispersant use decisions in spill response planning. *In* International Oil Spill Conference Proceedings: 1249–55.

Clayton, J.R., Payne, J.R., Farlow, J.S. (ed.). 1992. Oil spill dispersants mechanisms of action and laboratory tests. CK Smoley, Boca Raton, Florida, U.S.A. 88 p.

Coelho, G., Clark, J., and Aurand, D. 2013. Toxicity testing of dispersed oil requires adherence to standardized protocols to assess potential real world effects. Environ. Pollut. 177: 185–188.

CONCAWE. 1993. Ecotoxicological testing of petroleum products: Test methodology report No. 92/56. 22 p.

Cooney, J.D. 1995. Chapter 2, Freshwater tests. In: *Fundamentals of aquatic toxicology: effects, environmental fate, and risk assessment*. 2nd edition. Edited by Gary M. Rand. Washington DC, Taylor & Francis. 71–102.

Cotter, J. 2016. Husky Energy changes day it discovered Saskatchewan oil pipeline spill. (accessed July 30, 2016).

Couillard, C.M., Lee, K., Légaré, B., and King, T.L. 2005. Effect of dispersant on the composition of the water accommodated fraction of crude oil and its toxicity to larval marine fish. Environ. Toxicol. Chem. 24(6): 1496–1504.

Croce, B. and Stagg, R.M. 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 pigmented salmon syndrome. Environ. Toxicol. Chem. 16: 1921–1929.

Daaou, M. and Bendedouch, D. 2012. Water pH and surfactant addition effects on the stability of an Algerian crude oil emulsion. J. Saudi. Chem. Soc. 16: 333–337.

da Cruz, A.C.S., Leite, M.B.N., Rodrigues, L.E.A, and Nascimento, I.A. 2012. Estimation of biodiesel cytotoxicity by using acid phosphatase as a biomarker of lysosomal integrity. B. Environ. Contam. Tox. 89(2): 219–224.

Danion, M., Le Floch, S., Kanan, R., Lamour, F., and Quentel, C. 2011. Effects of in vivo chronic hydrocarbons pollution on sanitary status and immune system in sea bass (*Dicentarchus labrax* L.). Aquat. Toxicol. 105: 300–311.

Davis, J.C. 1977. Standardization and protocols of bioassays - Their role and significance for monitoring, research and regulatory useage (EPS-5-AR-77-1). *In* 3rd Aquatic Toxicity Workshop: 1–14.

Davison, W., Franklin, C.E., Mckenzie, J.C., and Dougan, M.C. 1992. The effects of acute exposure to the water soluble fraction of diesel fuel oil on survival and metabolic rate of an Antarctic fish (*Pagothenia borchgrevinki*). Comp. Biochem. Phys C. 102(1): 185–188.

Dew, W.A., Hontela, A., Rood, S.B., Pyle, G.G. 2015. Biological effects and toxicity of diluted bitumen (dilbit) and its constituents in freshwater systems. J. App. Toxicol. 35: 1219–1227.

Dietrich, D.E., Bowman, M.J., Korotenko, K.A., Hamish, M., and Bowman, E. 2014. Oil Spill Risk Management: Modeling Gulf of Mexico circulation and oil dispersal. Scrivener Publishing, Salem, Massachusetts.

Doe, K.G. and Wells, P.G. 1978. Acute aquatic toxicity and dispersing effectiveness of oil spill dispersants: Results of a Canadian Oil Dispersant Testing Program (1973 to 1977). In Chemical Dispersants for the Control of Oil Spills, ASTM STP 659: 50–65.

Doudoroff, P., Anderson, B.G., Burdick, G.E., Galtsoff, P.S., Hart, W.B., Patrick, R., Strong, E.R., and van Horn, W.M. 1951. Bioassay methods for the evaluation of acute toxicity of industrial wastes to fish. Sew. Ind. Waste. 23: 1380–1397.

Echols, B.S., Smith, A.J., Gardinali, P.R., and Rand, G.M. 2016a. The use of ephyrae of a scyphozoan jellyfish, *Aurelia aurita*, in the aquatic toxicological assessment of Macondo oils from the Deepwater Horizon incident. Chemosphere. 144: 1893–1900.

Echols, S.B., Smith, A., Gardinali, P., and Rand, G. 2016b. An evaluation of select test variables potentially affecting acute oil toxicity. Arch. Environ. Con. Toxicol. 70: 392–405.

Enbridge Energy. 2013. Enbridge Line 6B MP 608. Marshall, MI pipeline release. Supplement to the response plan for downstream impacted areas commonly referred to as the “Quantification of Submerged Oil Report”. Resubmitted version March 21, 2013. Prepared for the United States Environmental Protection Agency. 171 p.

ENG, 2011. Enbridge Northern Gateway, 2011, Northern Gateway Response to Federal Government IR No.1, ENG JRP Exhibit #B41-4, 246 p.

Environment Canada (EC). 1998. Biological test method; Toxicity tests using early life stages of salmonid fish (rainbow trout) EPS1/RM/28 Second Edition. Method Development and Applications Section (MDAS), Biological Methods Division, Environment Canada, Ottawa, ON.

Environment Canada (EC). 2007. Biological test method; Acute lethality test using rainbow trout EPS1/RM/09. Method Development and Applications Section (MDAS), Biological Methods Division, Environment Canada, Ottawa, ON.

Environment Canada (EC). 2011. Biological test method; Test of larval growth and survival using fathead minnows EPS1/RM/22 Second Edition. Method Development and Applications Section (MDAS), Biological Methods Division, Environment Canada, Ottawa, ON.

Environment Canada (EC). 2016. Oil properties database. Environmental Technology Centre. (accessed August 19, 2016).

Ernst, V.V., Neff, J.M., and Anderson, J.W. 1977. The effects of the water-soluble fractions of No. 2 fuel oil on the early development of the estuarine fish, *Fundulus grandis* Baird and Girard. Environ. Pollut. 14(1): 25–35.

Faksness, L.G., Brandvik, P.J., and Sydnes, L.K. 2008. Composition of the water accommodated fractions as a function of exposure times and temperatures. Mar. Pollut. Bull. 56: 1746–1754.

Federal On Scene Coordinator (FOSC). 2016. Desk Report for the Enbridge Line 6B Oil Spill Marshall, Michigan. April 2016. 241 p.

Fingas, M., Bier, I., Bobra, M., and Callaghan, S. 1991. Studies on the physical and chemical behavior of oil and dispersant mixtures. *In* International Oil Spill Conference Proceedings, 1: 419–426.

Forth, H.P., Mitchelmore, C.L., Morris, J.M., and Lipton, J. 2016. 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*. doi.org/10.1002/etc.3672.

Fragoso, N.M., Parrott, J.L., Hahn, M.E., and Hodson, P.V. 1998. Chronic Retene Exposure Causes Sustained Induction of Cyp1a Activity and Protein in Rainbow Trout (*Oncorhynchus Mykiss*). Environ. Toxicol. Chem. 17: 2347–53.

Freeman, L. 1953. A Standardized Method for Determining Toxicity of Pure Compounds to Fish. Sew. Ind. Waste. 25: 845–848.

Fucik, K.W., Carr, K.A., & Balcom, B.J. 1995. Toxicity of oil and dispersed oil to the eggs and larvae of seven marine fish and invertebrates from the Gulf of Mexico. In *The Use of Chemicals in Oil Spill Response*. ASTM International.

Gardiner, W.W., Word, J.Q., Word, J.D., Perkins, R.A., McFarlin, K.M., Hester, B.W., Word, L.S., and Ray, C.M. 2013. The acute toxicity of chemically and physically dispersed crude oil to key Arctic species under Arctic conditions during the open water season. Environ. Toxicol. Chem. 32(10): 2284–2300.

George-Ares, A., Clark, J.R., Biddinger, G.R., and Hinman, M.L. 1999. Comparison of test methods and early toxicity characterization for five dispersants. Ecotox. Environ. Safe. 42: 138–142.

George-Ares, A. and Clark, J.R. 2000. Aquatic toxicity of two Corexit® dispersants. Chemosphere. 40: 897–906.

Girling, A.E. 1989. Preparation of aqueous media for aquatic toxicity testing of oils and oil-based products: A review of the published literature. Chemosphere. 19: 1635–1641.

Girling, A.E., Whale, G.F., and Adema, D.M.M. 1994. A guideline supplement for determining the aquatic toxicity of poorly water-soluble complex mixtures using Water-Accommodated Fractions. Chemosphere. 29: 2645–2649.

González-Doncel, M., González, L., Fernández-Torija, C., Navas, J.M., & Tarazona, J.V., 2008. Toxic effects of an oil spill on fish early life stages may not be exclusively associated to PAHs: studies with Prestige oil and medaka (*Oryzias latipes*). Aquat. Toxicol. 87(4): 280–288.

Government of Canada (GOC). 2013. Properties, composition and marine spill behaviour, fate and transport of two diluted bitumen products from the Canadian oil sands. Federal Government Technical Report. 1–85 p.

Hanson, M.L., Wolff, B.A., Green, J.W., Kivi, M., Panter, G.H., Warne, M.St.J., Ågerstrand, M., and Sumpter, J.P. 2016. How we can make ecotoxicology more valuable to environmental protection. Sci. Total Environ. doi:10.1016/j.scitotenv.2016.07.160.

Harris, C.A., Scott, A.P., Johnson, A.C., Panter, G.H., Sheahan, D., Roberts, M., and Sumpter, J.P. 2014. Principles of sound ecotoxicology. Ecotox. Environ. Sci. Tech. 48: 3100–3111.

Hedayati, A. and Jahanbakhshi, A. 2012. The effect of water-soluble fraction of diesel oil on some hematological indices in the great sturgeon *Huso huso*. Fish Physiol. Biochem. 38(6): 1753–1758.

Hedtke, S.F. and Puglisi, F.A.1982. Short-term toxicity of five oils to four freshwater species. Arch. Environ. Con. Tox. 11(4): 425–430.

Heintz, R.A., Short, J.W., and Rice, S.D. 1999. Sensitivity of fish embryos to weathered crude oil: Part II. Increased mortality of pink salmon (*Oncorhynchus gorbuscha*) embryos incubating downstream from weathered Exxon Valdez crude oil. Environ. Toxicol. Chem. 18: 494–503.

Heintz, R.A., Rice, S.D., Wertheimer, A.C., Bradshaw, R.F., Thrower, F.P., Joyce, J.E., and Short, J.W. 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.

Hemmer, M.J., Barron, M.G., and Greene, R.M. 2011. Comparative toxicity of eight oil dispersants, Louisiana sweet crude oil (LSC), and chemically dispersed LSC to two aquatic test species. Environ. Toxicol. Chem. 30(10): 2244–2252.

Hemmingsen, P.V., Silset, A., Hannisdal, A., and Sjöblom, J. 2005. Emulsions of heavy crude oils. I: Influence of viscosity, temperature, and dilution. J. Dispers. Sci. Tech. 26: 615–627.

Hodson P.V., Khan, C., Saravanabhavan, G., Clarke, L., Brown, R.S., Hollebone, B.,Wang, Z., Short, J., Lee, K., and King, T. 2007. AlkylPAH in crude oil cause chronic toxicity to early life stages of fish. *In* Proceedings of 28th Arctic and Marine Oilspill Program Technical Seminar (AMOP). 291–300.

Hollebone, B.P. 2015. Appendix A. The oil properties data appendix. *In*: Handbook of Oil Spill Science and Technology. Edited by M.F. Fingas. Wiley & Sons, Inc., Hoboken, NJ. pp. 577–681.

Hook, S.E., Lampi, M.A., Febbo, E.J., Ward, J.A., and Parkerton, T.F. 2010. Temporal patterns in the transcriptomic response of rainbow trout, *Oncorhynchus mykiss*, to crude oil. Aquat. Toxicol. 99: 320–329.

Incardona, J.P., Carls, M.G., Day, H.L., Sloan, C.A., Bolton, J.L., Collier, T.K., and Schoiz, N.L. 2009. Cardiac arrhythmia is the primary response of embryonic Pacific Herring (*Clupea Pallasi)* exposed to crude oil during weathering. Environ. Sci. Tech. 43: 201–207.

Incardona, J.P., Swarts, T.L., Edmunds, R.C., Linbo, T.L., Aquilina-Beck, A., Sloan, C.A., Gardner, L.D., Block, B.A., and Scholz, N.L. 2013. Exxon Valdez to Deepwater Horizon: Comparable toxicity of both crude oils to fish early life stages. Aqua. Toxicol. 142: 303–316.

Kaku, V.J., Boufadel, M.C., and Venosa, A.D. 2006. Evaluation of turbulence parameters in laboratory flasks used for dispersant effectiveness testing. J. Environ. Eng. 132: 93–101.

Kennedy, C.J. and Farrell, A.P. 2005. Ion homeostasis and interrenal stress responses in juvenile Pacific herring, *Clupea pallasi*, exposed to the water-soluble fraction of crude oil. J. Exp. Mar. Biol. Ecol. 323: 43–56.

Khelifa, A., Fingas, M., Hollebone, B.P., and Brown, C.E. 2007. Effects of chemical dispersants on oil physical properties and dispersion. *In* 13th Arctic and Marine Oil Spill Program Technical Seminar (AMOP), Environment Canada, Ottawa, Ontario, 105–116.

King, T.L., Robinson, B., Boufadel, M., and Lee, K. 2014. Flume tank studies to elucidate the fate and behavior of diluted bitumen spilled at sea. Mar Pollut. Bull. 83(1): 32–37. http://doi.org/10.1016/j.marpolbul.2014.04.04.

Kiparissis, Y., Akhtar, P., Hodson, P.V., and Brown, R.S. 2003. Partition Controlled Delivery (PCD) of toxicants: a novel *in vivo* approach for embryotoxicity testing. Environ. Sci. Technol. 37: 2262–2266.

Kuhl, A.J., Nyman, J.A., Kaller, M.D., and Green, C.C. 2013. Dispersant and salinity effects on weathering and acute toxicity of South Louisiana crude oil. Environ. Toxicol. Chem. 32: 2611–2620.

Landrum, P.F., Chapman, P.M., Neff, J., and Page, D.S. 2013. Influence of exposure and toxicokinetics on measures of aquatic toxicity for organic contaminants: A case study review. Integrat. Environ. Assess. Manag. 9999: 1–15.

LaRoche, G., Eisler, R., and Tarzwell, C.M. 1970. Bioassay procedures for oil and oil dispersant toxicity evaluation. J. Water. Pollut. Cont. Fed. 42: 1982–1989.

Lee, R.F. 2003. Photo-oxidation and photo-toxicity of crude and refined oils. Spill Sci. Technol. B. 8 (2): 157–162.

Lee, K., Boufadel, M., Chen, B., Foght, J., Hodson, P.V., Swanson, S., and 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

Lessard, R.R. and Demarco, G. 2000. The significance of oil spill dispersants. Spill. Sci. and Tech. Bull. 6: 59–68.

Little, E.E., Cleveland, L., Calfee, R., and Barron, M.G. 2000. Assessment of the photoenhanced toxicity of a weathered oil to the tidewater silverside. Environ. Toxicol. Chem. 19(4): 926–932.

Lunel, T. 1993. Dispersion: oil drop size measurements at sea. In Proceedings, 19th Arctic Marine Oilspill Program (AMOP) Technical Seminar, Environment Canada, Ottawa, Ontario, 1023–1105.

Madison, B.N., Hodson, P.V., and Langlois, V.S. 2015. Diluted bitumen causes deformities and molecular responses indicative of oxidative stress in Japanese Medaka embryos. Aqua. Toxicol. 165: 222–230.

Maher, W.A. 1982. Preparation and characterization of water-soluble fractions of crude and refined oils for use in toxicity studies. Bull. Environ. Contamin. Toxicol. 29: 268–272.

Maher, W.A. 1986. Preparation of water soluble fractions of crude oils for toxicity studies. Bull. Environ. Contamin. Toxicol. 36: 226–229.

MARPOL. 2005. Regulation 21 - Prevention of oil pollution from oil tankers carrying heavy grade oil as cargo.

Martin, J.D., Adams, J., Hollebone, B.P., King, T., Brown, R.S., and Hodson, P.V. 2014. Chronic toxicity of heavy fuel oils to fish embryos using multiple exposure scenarios. Environ. Toxicol. Chem. 33: 677–687.

Marty, G.D., Heintz, R.A., and Hinton, D.E. 1997. Histology and teratology of pink salmon larvae near the time of emergence from gravel substrate in the laboratory. Can. J. Zool. 75: 978–988.

Mattos, J.J., Siebert, M.N., Luchmann, K.H., Granucci, N., Dorrington,T., Stoco, P.H., Gisard, E.C., and Bainy, A.C. 2010. Differential gene expression in *Poecillia vivpara* exposed to diesel oil water accommodated fraction. Mar. Environ. Res. 69: S31–S33.

McIntosh, S., King, T., Wu, D., and Hodson, P.V. 2010. Toxicity of dispersed weathered crude oil to early life stages of Atlantic herring (*Clupea harengus*). Environ. Toxicol. Chem. 29: 1160–1167.

McSheffrey, E. 2016. Environment Canada opens investigation as blobs of Husky oil contaminate river. National Observer. (accessed July 30, 2016).

Melbye, A.G., Brakstad, O.G., Hokstad, J.N., Gregersen, I.K., Hansen, B.H., Booth, A.M., Rowland, S.J., and Tollefsen, K.E. 2009. Chemical and toxicological characterization of an unresolved complex mixture-rich biodegraded crude oil. Environ. Toxicol. Chem. 28: 1815–1824.

Moles, A. and Rice, S.D. 1983. Effects of crude oil and naphthalene on growth, caloric content, and fat content of pink salmon juveniles in seawater. T. Am. Fish Soc. 112(2A): 205–11.

National Academies of Sciences, Engineering, and Medicine (NAS). 2016. Spills of Diluted Bitumen from Pipelines: A Comparative Study of Environmental Fate, Effects, and Response. Washington, DC: The National Academies Press.

National Ground Water Association. 2010. Brackish Groundwater. NGWA Information Brief. 7 pages.

National Oceanic and Atmospheric Administration (NOAA). 2008. Estuaries. NOAA ocean service education. (accessed August 23, 2016).

National Research Council (NRC). 2005. Toxicological effects of dispersants and dispersed oil. In: *Oil Spill Dispersants. Efficacy and Effects.* National Academies Press, Washington, DC. pp. 193–275.

Nexen. 2015. Long Lake Pipeline Failure, Incident Overview.

Nickel, R. and Williams, N. 2016. Husky says Canadian oil spill started day before pipeline was shut. Reuters Canada. (accessed July 30, 2016).

Nordtug, T., Olsen, A.J., Altin, D., Meier, S., Overrein, I., Hansen, B.H., and Johansen, O. 2011. Method for generating parameterized ecotoxicity data of dispersed oil for use in environmental modelling. Mar. Pollut. Bull. 62: 2106–2113.

Østgaard, K. and Jensen, A. 1983. Preparation of aqueous petroleum solutions for toxicity testing. Environ. Sci. Tech. 17: 548–553.

Philibert, D.A., Philibert, C.P., Lewis, C., and Tierney, K. 2016. A comparison of diluted bitumen (dilbit) and conventional crude oil toxicity to developing zebrafish. Environ. Sci. Tech. 50: 6091–6098.

POLARIS Applied Sciences (POLARIS). 2013. A comparison of the properties of diluted bitumen crudes with other oils.

Prystupa, M. 2015. Massive Alberta oil spill bigger than Kalamazoo disaster. National Observer. (accessed 12 June 2016).

Ramachandran, S.D., Hodson, P.V., Khan, C.W., and Lee, K. 2004. Oil dispersant increases PAH uptake by fish exposed to crude oil. Ecotox. Environ. Safe. 59: 300–308.

Ramachandran, S.D., Sweezey, M.J., Hodson, P.V., Boudreau, M., Courtenay, S.C., Lee, K., King, T., and Dixon, J.A. 2006. Influence of salinity and fish species on PAH uptake from dispersed crude oil. Mar. Pollut. Bull. 52: 1182–1189.

Rand, G.M., Wells, P.G., and McCarty, L.S. 1995. Chapter 1, Introduction to Aquatic Toxicology In: *Fundamentals of aquatic toxicology: effects, environmental fate, and risk assessment*. 2nd edition. Edited by Gary M. Rand. Washington DC, Taylor & Francis. 3–67.

Redman, A.D., McGrath J.A., Stubblefield, W.A., Maki, A.W., and Di Toro, D.M. 2012. Quantifying the concentration of crude oil microdroplets in oil-water preparations. Environ. Toxicol. Chem. 31: 1814–1822.

Redman, A.D., Parkerton, T.F., Letinski, D.J. Manning, R.G., Adams, J.E., and Hodson, P.V. 2014. Evaluating toxicity of heavy fuel oil fractions using complementary modeling and biomimetic extraction methods. Environ. Toxicol. Chem. 33: 2094–2104.

Redman, A.D. and Parkerton, T.F. 2015. Guidance for improving comparability and relevance of oil toxicity tests. Mar. Pollut. Bull. 98: 156–170.

Rico-Martínez, R., Snell, T.W., and Shearer, T.L. 2013. Synergistic Toxicity of Macondo crude oil and dispersant Corexit 9500A to the *Brachionus plicatilis* species complex (*Rotifera*). Environ. Pollut. 173: 5–10.

Rhoton, S.L., Perkins, R.A., Braddock, J.F, and Behr-Andres, C. 2001. A cold-weather species' response to chemically dispersed fresh and weathered Alaska North Slope crude oil. *In* International Oil Spill Conference 2001: 1231–36.

Rufli, H., Fisk, P.R., Girling, A.E., King, J.M.H., Länge, R., Lejeune, X., Stelter, N., Stevens, C., Suteau, P., Tapp., J., Thus, J., Versteeg, D.J., and Neissen, H.J. 1998. Aquatic toxicity testing of sparingly soluble, volatile, and unstable substances and interpretation and use of data. Ecotox. Environ. Safe. 39: 72–77.

Santos, C.A., Lenz, D., Brandao, G.P., Chippari-Gomes, A.R., and Gomes, L.C. 2013. Acute toxicity of the water-soluble fraction of diesel in *Prochilodus vimboides* Kner (Characiformes: Prochilodontidae). Neotrop. Icthyol. 11: 193–198.

Schein, A., Scott, J.A., Mos, L., and Hodson, P.V. 2009. Oil dispersion increases the apparent bioavailibilty and toxicity of diesel to rainbow trout (*Oncorhynchus mykiss*). Environ. Toxicol. Chem. 28: 595–602.

Shiu, W.Y., Bobra, M., Bobra, A.M., Maijanen, A., Suntio, L., and Mackay, D. 1990. The water solubility of crude oils and petroleum products. Oil & Chemical Pollution 7: 1–2.

Shukla, P., Gopalani, M., Ramteke, D.S., and Wate, S.R. 2007. Influence of salinity on PAH uptake from water soluble fraction of crude oil in *Tilapia mossambica*. B. Environ. Contam. Tox. 79: 601–605.

Simonato, J.D., Guedes, C.L., and Martinez, C.B. 2008. Biochemical, physiological, and histological changes in the neotropical fish *Prochilodus lineatus* exposed to diesel oil. Ecotox. Environ. Safe. 69(1): 112–120.

Singer, M.M., Aurand, D., Bragin, G.E., Clark, J.R., Coelho, G.M., Sowby, M.L., and 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.

Singer, M.M., Aurand, D.V., Coelho, G.M., Bragin, G.E., Clark, J.R., Sowby, M., and Tjeerdema, R.S. 2001a. Making, measuring, and using water-accommodated fractions of petroleum for toxicity testing. *In* International Oil Spill Conference 2001 2: 1269–1274.

Singer, M.M., Jacobson, S., Tjeerdema, R.S., and Sowby, M. 2001b. Acute effects of fresh versus weathered oil to marine organisms: California findings. *In* International Oil Spill Conference Proceedings 1: 1263–1268.

SL Ross. 2012. Meso-scale weathering of Cold Lake bitumen/condensate blend. SL Ross Environmental Research Ltd, #200 - 1140 Morrison Drive, Ottawa, ON K2H 8S9

Sørhus, E., Edvardsen, R.B., Karlsen, O., Nordtug, T., van der Meeren, T., Thorsen, A., Harman, C., Jentoft, S., and Meier, S. 2015. Unexpected interaction with dispersed crude oil droplets drives severe toxicity in Atlantic haddock embryos. Plos One. 10(4): e0124376.

Speight, J.G. 2006. The chemistry and technology of petroleum. CRC Press, New York, NY. 143-174, 4th Edition.

Speight, J.G. 2014. The Chemistry and Technology of Petroleum. 5th Edition. CRC Press, Hoboken, NJ.

Sprague, J.B. 1969. Measurement of Pollutant Toxicity to Fish I. Bioassay methods for acute toxicity. Water. Res. 3: 793–821.

Sprague, J.B. 1970. Measurement of Pollutant Toxicity to Fish. II. Utilizing and applying bioassay results. Water. Res. 4: 3–32.

Sprague, J.B. 1971. Measurement of pollutant toxicity to fish - III Sublethal effects and ‘safe’ concentrations. Water. Res. 5: 245–266.

Sprague, J.B. 1995. Appendix C, Factors that Modify Toxicity. In: *Fundamentals of aquatic toxicology: effects, environmental fate, and risk assessment*. 2nd edition. Edited by Gary M. Rand. Washington DC, Taylor & Francis. 1012–1051.

Stokes, V.K. and Harvey, A.C. 1973. Drop size distributions in oil water mixtures. *In* International Oil Spill Conference Proceedings. 1973: 457–465.

Swedmark, M., Granmo, Å., and Kollberg, S. 1973. Effects of oil dispersants and oil emulsions on marine animals. Water. Res. 7(11): 1649–1672.

Timme-Laragy, A.R., Cockman, C.J., Matson, C.W., and Di Giulio, R.T. 2007. Synergistic Induction of AHR regulated genes in developmental toxicity from co-exposure to two model PAHs in zebrafish. Aqua. Toxicol. 85: 241–250.

Tjeerdema, R.S., Lin, C.Y., Anderson, B.S., Viant, M.R., Crane, D., and Sowby, M.L. 2007. Effects of oil on the smolts of salmon (*Oncorhynchus tshawytscha*)–preliminary muscle results. Proceedings: Papers, p.207.

Tsvetnenko, Y. and Evans, L. 2002. Improved approaches to ecotoxicity testing of petroleum products. Mar. Pollut. Bull. 45: 148–156.

USEPA. Method EPA SW-846 (accessed March 1, 2017)

Ward, G.S. 1995. Chapter 3, Saltwater Tests. In: *Fundamentals of aquatic toxicology: effects, environmental fate, and risk assessment*. 2nd edition. Edited by Gary M. Rand. Washington DC, Taylor & Francis. 103–133.

Wilson, K.W. 1977. Acute toxicity of oil dispersants to marine fish larvae. Mar. Biol. 40: 65–74.

Woodward, D.F., Little, E.E., and Smith, L.M. 1987. Toxicity of five shale oils to fish and aquatic invertebrates. Arch. Environ. Con. Tox. 16(2): 239–246.

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

Yang, C., Wang, Z., Yang, Z., Hollebone, B., Brown, C.E., Landriault, M., and Fieldhouse, B. 2011. Chemical fingerprints of Alberta oil sands bitumen and related petroleum products. Environ. Forensics. 12: 173–188.

Yoon, S., Bhatt, S.D., Lee, W., Lee, H.Y., Jeong, S.Y., Baeg, J.-O., and Lee, C.W. 2009. Separation and characterization of bitumen from Athabasca oil sand. Korean. J. Chem. Eng. 26: 64–71.

Yu, X., Xu, C., Liu, H., Xing, B., Chen L., and Zhang, G. (2015). Effects of crude oil and dispersed crude oil on the critical swimming speed of puffer fish, *Takifugu rubripes.* B. Environ. Contam. Tox. 94: 549–553.

Ziolli, R.L. and Jardim, W.F. 2002. Operational problems related to the preparation of the seawater soluble fraction of crude oil. J. Environ. Monit. 4: 138–141.

9 APPENDICES

APPENDIX 1: CROSERF PROTOCOLS

Table A1: CROSERF protocols for marine oil acute toxicity tests with Water Accommodated Fractions (WAF) and Chemically Enhanced-WAF (CEWAF). This table was modified from Aurand and Coelho 2005[a] and the information was compiled from Aurand and Coelho 2005[a] and Barron and Ka'aihue 2003[b].

<table>
<tr><th colspan="2">PARAMETER</th><th>WAF</th><th>CEWAF</th></tr>
<tr><td colspan="2">Water Preparation</td><td colspan="2">• Minimum filtration 0.45 µm
• Local sea water preferred; dilute with deionized water as necessary (reconstituted sea water acceptable)
• If natural sea water is unavailable, use water known to support good survival and health of test organisms
• Use the same water for preparing WAF/CEWAF as the dilution water</td></tr>
<tr><td rowspan="2">Oil Preparation</td><td><i>Fresh</i></td><td colspan="2">• Include reference oil as a positive control; PBCO (Prudhoe bay Crude Oil)
• No preparation necessary
• See storage and handling</td></tr>
<tr><td><i>Weathered</i></td><td colspan="2">• Include reference oil as a positive control; PBCO
• Should be artificially weathered under controlled conditions (avoid/minimize irreproducible natural weathering)
• Target equivalent to 6 to 24 hours of weathering (~20% for crude oils)
• Follow ASTM Method D2892 (D86 is an acceptable alternative)
• Recommend fractional distillation by a commercial analytical lab to characterize weathered oil
• See storage and handling</td></tr>
<tr><td colspan="2">Mixing Vessel Size</td><td colspan="2">• Based on volume of WAF/CEWAF required
• 1 L to 20 L</td></tr>
<tr><td colspan="2">Stir-Bar Size</td><td colspan="2">• Should be selected to match container size and/or stir plate
• Recommended ~1 inch for 2 L and ~2 inches for 20 L</td></tr>
<tr><td colspan="2">Addition of Oil</td><td colspan="2">• Add known volume of oil at centre of container using gas-tight Hamilton syringe during mixing
• Calculate delivery mass by difference
• 0.01-25 g/L loading of oil in water [b] (no serial dilution)</td></tr>
<tr><td colspan="2">Addition of Dispersant</td><td>N/A</td><td>• Corexit 9500 and/or 9527, 1:10 DOR [b]
• Add known volume of dispersant to oil (sequentially) at centre of vortex using gas-tight Hamilton syringe after the vortex has been established
• Calculate delivery mass by difference
• Do not pre-mix oil and dispersant</td></tr>
<tr><td colspan="2">Headspace</td><td colspan="2">• 20-25% headspace when mixing container is filled to base of shoulder</td></tr>
</table>

<table>
<tr><th colspan="2">PARAMETER</th><th>WAF</th><th>CEWAF</th></tr>
<tr><td colspan="2">**Mixing Energy**</td><td>• Originally 3-4 rps (180-240 rpm), or as low as necessary to avoid vortex (avoid settling and emulsification)</td><td>• Vortex 20-25% of water depth
• Conduct a pre-test on oil to determine how the oil and dispersant should be added to the water</td></tr>
<tr><td rowspan="2">**Mixing Conditions**</td><td>*Temperature*</td><td colspan="2">• Controlled temperature environment
• Temperature should be appropriate to the test organism</td></tr>
<tr><td>*Light*</td><td colspan="2">• Mix in darkness [a] or in laboratory lighting (fluorescent) [b]</td></tr>
<tr><td colspan="2">**Mixing Duration**</td><td>• Based on the oil type and loading rate
• Stability of solutions determined by chemistry</td><td>• 18 hours recommended, but should be optimized for each oil, like WAF
• Should not exceed 24 hours (i.e., to avoid biodegradation)
• Stability of solutions determined by chemistry</td></tr>
<tr><td colspan="2">**Settling Duration**</td><td>• No settling prior to extraction of aqueous layer
• Recommended use immediately after mixing
• If necessary, WAF can be stored a maximum of 24 hours prior to use</td><td>• Settle 3 hours minimum, maximum 6 hours
• Use immediately after settling</td></tr>
<tr><td rowspan="3">**Storage and Handling**</td><td>*Oil*</td><td colspan="2">• Store in tightly sealed glass or metal containers with minimal headspace
• Store in darkness at < 5 °C
• Recommend chemical characterization of stored oil to monitor changes in composition over time</td></tr>
<tr><td rowspan="2">*Solution*</td><td>• Can be stored in Teflon bags</td><td>• Should be used immediately (but appears stable for at least 24 hours)</td></tr>
<tr><td colspan="2">• Do not freeze - waxes and asphaltenes may separate from oil
• Samples for chemical analysis should be immediately acidified and analyzed within 7 (BTEX) to 14 days (semi-volatiles)</td></tr>
<tr><td rowspan="3">**Toxicity Tests**</td><td>*Test Organisms*</td><td colspan="2">• Guidelines for animal aquaculture/handling should follow those published by ASTM, OECD, EPA, etc.
• Introduce organisms to treatments randomly
• Acclimation procedures, holding tank salinity and temperature, photo-period and light intensity, feed, and water quality parameters should be appropriate for the test organism
• Observe mortality and other effects daily
• Do not remove dead animals</td></tr>
<tr><td>*Treatments*</td><td colspan="2">• Serial dilution not recommended
• Use variable loading instead</td></tr>
<tr><td>*Exposure Regime*</td><td colspan="2">Spiked Exposure:
• Test initiated by flushing all chambers with clean, aerated, filtered water, then filling with test solution
• Flow-through rate 2 mL/min (in 240 mL CROSERF Flow-Through Toxicity Test Chamber)</td></tr>
</table>

PARAMETER		WAF	CEWAF
		• Chamber sealed after introduction of animals • Monitor decline in concentration for the first 6 to 8 hours • Average exposure of 96 hours (acute)	
		Constant Exposure: • Test initiated by filling test containers with clean, aerated, filtered water, then filling with test solution • Aeration 1-2 bubbles/s • Containers covered with glass after introduction of animals • Remove 90% of old solution every 24 h and replace with fresh solution • Average exposure of 96 hours (acute)	
	Endpoints	• Lethality (i.e., 96 h LC50) • “Initial effect” narcosis/moribundity (0.5 to 1h EC50)	
Chemical Characterization	• Measure TPH and volatile concentrations in ppm		
	TPH ($C_{10} – C_{36}$)	• Measured in fresh WAF and CEWAF solutions prior to toxicity testing and at the end of experiments for static tests; prior to each renewal for constant exposures; at one or more points during the experiment for spiked exposures • Use GC-FID. If using another method, provide a comparison of the method used to GC-FID • Use at least one surrogate standard: o-terphenyl with 70-120% recovery • Use one internal standard that does not co-elute with specific oil • Do not correct for recovery • Detection limit ≤10 ppb for any specific n-alkane in the standard • Measure PAHs (optional)	
	Volatiles ($C_6 < C_{10}$)	• Measured in initial oils, fresh WAF/CEWAF prior to toxicity testing, at the end of the experiment for static tests; prior to each renewal for constant exposures; at one or more points during the experiment for spiked exposures • Use GC-FID or GC-MS. If using another method, compare the results to GC-FID or GC-MS • Use an internal standard • D-toluene recommended as surrogate standard if using GC-MS • See Appendix 2, Table 2A for list of minimum target analytes	
	Compound Specific Semi-Volatiles	• Optional, recognizing cost and time limitations • Choose a specific analytical method based on target analytes, available equipment, and performance • See Appendix 2, Table 2A for list of optional target analytes	
	Dispersant-only Solutions	N/A	• Use a series of concentrations (1 – 1000 ppm) of dispersant to create calibration curve • Use same sea water diluent as for toxicity tests • Analyze samples within 2 hours by UV-VIS spectrophotometry • Use absorption maxima for dispersant for all analyses

PARAMETER	WAF	CEWAF
Statistical Analysis	• Use Probit test if data are normally distributed • Use Spearman-Karber or Trimmed Spearman-Karber test for non-parametric distributions (if there is no control mortality) • Use the binomial method in situations where there are no partial kills (i.e., 0% or 100% mortality), or re-run the test using a different range of concentrations • Report statistical methods whenever presenting LC50, EC50, or other toxicity data	

APPENDIX 2: TARGET ANALYTE LISTS FOR VOLATILE AND SEMI-VOLATILE ANALYSIS

Table A2.1: The minimum target analyte lists for volatile and semi-volatile analysis by CROSERF was described in Aurand & Coelho (2005; Table 3.3 and Table 3.4).

Analytical method	**Minimum Analyte List**	
Volatile analysis	Saturates	
	• 2-methylpentane • Hexane • Cyclopentane • 2,4-diemthylpentane • Cyclohexane	• Heptane • Cycloheptane • Octane • Nonane
	Unsaturates	
	• Benzene • Toluene • Ethylbenzene • p-xylene	• m-xylene • o-xylene • n-propylbenzene • C_3-benzenes
Semi-volatile analysis (optional)	• Naphthalenes (C_0-C_4) • Biphenyl • Fluorenes (C_0-C_3) • Dibenzothiophenes (C_0-C_4) • Phenanthrenes (C_0-C_4) • Fluoranthene • Pyrenes (C_0-C_4) • Chrysenes (C_0-C_4) • Benzo(a)anthracene	• Benzo(b)fluoranthene • Benzo(k)fluoranthene • Benzo(e)pyrene • Benzo(a)pyrene • Perylene • Indeno(1,2,3-cd)pyrene • Dibenzo(a,h)anthracene • Benzo(g,h,i)perylene

Table A2.2: Recommended target analyte lists for volatile and semi-volatile analysis based on this review. Unsaturates TPAH51 list adapted from TPAH50 list from Forth et al. (2016) plus perylene from Aurand and Coelho (2005).

Analytical method	**Minimum Analyte List**	
Volatile analysis	Saturates	
	• Pentane • 2-methylpentane • Hexane • Cyclopentane • 2,4-diemthylpentane	• Cyclohexane • Heptane • Cycloheptane • Octane • Nonane
	Unsaturates	
	• Benzene • Toluene • Ethylbenzene • p-xylene • m-xylene	• o-xylene • n-propylbenzene • Cumene • other C_3-benzenes
Semi-volatile analysis (optional)	• Naphthalenes (C_0-C_4) • Biphenyl • Dibenzofuran • Acenaphthylene • Acenaphthene • Fluorenes (C_0-C_3) • Anthracene • Phenanthrene • Phenanthrenes (C_1-C_4)/ Anthracenes (C_1-C_4) • Dibenzothiophenes (C_0-C_4 • Benzo(b)fluorene • Fluoranthene • Pyrene • Pyrenes (C_1-C_4)/ Fluoranthenes (C_1-C_4)	• Naphthobenzothiophenes (C_0-C_4) • Benzo(a)anthracene • Chrysene + Triphenylene • Chrysenes (C_1-C_4) • Benzo(b)fluoranthene • Benzo(j+k)fluoranthene • Benzo(a)fluoranthene • Benzo(e)pyrene • Benzo(a)pyrene • Perylene • Indeno(1,2,3-cd)pyrene • Dibenzo(a,h)anthracene • Benzo(g,h,i)perylene

APPENDIX 3: CATEGORIZATION OF FISH SPECIES USED IN OIL TOXICITY TESTS BY HABITAT

Table A3: Categorization of fish species used in oil toxicity tests by habitat (salt water, freshwater, or both) and preference of water temperature (warm water or cold water). Entries for each fish species include the Latin names (Genus species). Fish found in Canadian waters are underlined (Department of Fisheries and Oceans; Aquatic Species). Species with an asterisk are ECCC Standard Test Species[1]*.*

	Salt Water Species	Freshwater Species	Both (Anadromous Species)
Warm Water Species	• Brown spotted grouper (*Ephinephelus chlorostignma*) • Cockscomb prickleback (*Anoplarchus purpurescens*) • European sea bass (*Dicentrarchus labrax*) • Flounder (*Pleuronectes flesus*) • Gulf killfish (*Fundulus grandis*) • Grumecha (*Prochilodus vimboides)* • Inland silverside minnow (*Menidia beryllina*) • Longnose killifish (*Fundulus similis)* • Mahi-mahi (*Coryphaena hippurus*) • Marine medaka (*Oryzias melastigma*) • Olive flounder (*Paralichthys olivaceus*) • Rabbit fish (*Siganus canaliculatus*) • Red drum (*Sciaenops ocellatus*) • Red sea bream (*Pagrus major*) • Red snapper (*Lujanus argentimaculatus*) • Rockfish (*Sebastes schlegeli*) • Sheepshead minnow (*Cyprinodon variegatue*) • Snapper (*Chrysophrys auratus*) • Spot (*Leiostomus xanthurus*) • Spotted seatrout (*Cynoscion nebulosus*) • Spotted seabass (*Lateolabrax maculates*) • Spotted wrasse (*Notolabrus celidotus)* • Topsmelt (*Atherinops affinis*) • Yellowtail kingfish (*Seriola lalandi*)	• African catfish (*Clarias gariepinus*) • Amazonian cichlid (*Astronatus ocellatus*) • American flagfish (*Jordanella floridae*) • *Astyanax* sp. • Crimson-spotted rainbowfish (*Melanotaenia fluviatilis*) • Guppy (*Poecilia vivipara*) • Japanese medaka (*Oryzias latipes*) • Sablo (*Prochilodus lineatus)* • Tilapia (*Oreochromis niloticus*) • Tilapia (*Tilapia mossambica*) • Tropical fish tambaqui (*Colossoma macropomum*) • Zebrafish (*Danio rerio*)	• Australian bass (*Macquaria novemaculeata*) • Colorado squawfish (*Ptychocheilus lucius)* • Great sturgeon (*Huso huso*) • Mummichog (*Fundulus heteroclitus*)

---

[1] The data were compiled from: Fish Base, Fundulus heteroclitus Assessment, DFO's Haddock Fact Sheet, Cod Fact Sheet (Norway), and Animal World Pet Encyclopedias.

	Salt Water Species	Freshwater Species	Both (Anadromous Species)
Cold Water Species	• Arctic cod (*Boreogadus saida)* • Atlantic cod (*Gadus morhua)* • Atlantic haddock (*Melanogrammus aeglefinus*) • Atlantic herring (*Clupea harengus*) • Atlantic menhaden (*Brevoortia tyrannus*) • Atlantic silverside (*Menidia menidia*) • Bald notothen (*Pagothenia borchgrevinki)* • Capelin (*Mallotus villosus*) • Crescent gunnel (*Pholis laeta*) • Cunner (*Tautogolabrus adspersus)* • Dolly varden (*Salvelinus malma*) • Golden grey mullet (*Liza aurata*) • Hogchoker (*Trinectes maculatus)* • Longhorn sculpin *(Myoxocephalus octodecemspinosus)* • Marine pejerrey (*Odontesthes argentinensis*) • Mullet (*Mugil cephalus*) • North East Arctic cod (*Gadus morhua*) • Pacific herring (*Clupea pallasi*) • Puffer fish (*Takifugu rubripes*) • Sculpin *(Myoxocephalus sp., polyacanthocephalus)* • Starry flounder (*Platichthys stellatus*) • Thinlipped grey mullet (*Liza ramada*) • Tubesnouts (*Aulorhynchus flavidus*) • Turbot (*Scophthalmus maximus L.*) • Walleye polluck (*Theragra chalcogrammus*)	• Caspian roaches (*Rutilus caspicus*) • Catfish (*Hoplosternum littorale, bidorsalis*) • Fathead minnow (*Pimephales promelas*)*	• Atlantic salmon (*Salmo salar*) • Chinook salmon (*Onchorhyncus tshawytsch*) • Cutthroat trout (*Salrno clarki*) • Pacific salmon (*Oncorhynchus nerka)* • Pink salmon (*Oncorhynchus gorbuscha*) • Rainbow trout (*Oncorhynchus mykiss*)*

APPENDIX 4: RECOMMENDED REPORTING REQUIREMENTS FOR TOXICITY TESTS, ADAPTED FROM HANSON ET AL. (2016)

Table A4: Additional guidance from other sources or specific to a test oil was added with the reference noted in superscript; CONCAWE 1993[a], Redman & Parkerton 2015[b], additions from this paper[c]. When sections are referred to in the text from Hanson et al. (2016) the section was corrected to match the numbering in this table, rather than the numbering in the publication.

Minimum reporting requirements for ecotoxicology studies

1. Test compound source and properties

Source and purity provided?

- Source, purity, and composition of the test substance; specifically, percent active ingredient including levels or ratios of components and isomers and impurities.
    a. Include the description of the oil (i.e. crude oil, refined petroleum products) and description of the weathering state[b]
    b. Oil chemistry[c]
    c. For chemical dispersions provide the chemistry and characteristics[c]

Technical name?

- Technical name (e.g., International Union of Pure and Applied Chemistry (IUPAC) or registration no., Chemical Abstract Service number (CAS), batch number) and formulated product, brand or trade names.
    a. Information about the extraction, refinement, storage and handling of oil[c]

2. Experimental design

Hypotheses, if any, stated?

- Hypotheses and objectives of the study should be clearly stated, even if the hypothesis is as simple as ‘compound X causes a 50% reduction in egg-laying relative to control at a concentration less than its aqueous solubility’.

Number of treatments and their exposure levels?

- The number of treatments and the nominal concentrations.

Number and type of controls?

- The number and types of controls (e.g., positive, negative, or solvent). If a solvent carrier is used then its concentration in each treatment and control should be equal and stated.

Duration of exposures?

- The methods for creating and storing the stock and working solutions and the duration of storage.
    a. Description of the methods for creating solutions for oil toxicity testing recommended[a]:
        - type of medium prepared (WAF or Dispersion)
            i. & Oiled substrate, e.g. gravel, Nozzles that deliver flor of oil droplets[b]
            ii. Dispersion procedure: physical (low or high mixing energy), chemical (type of dispersant, DOR), duration of equilibrium and settling periods, and temperature and salinity of aqueous media[b]
        - type of mixing system
        - mass of test material and medium (oil-to-water ratio, oil-to-substrate ratio[c])
        - geometry of mixing vessel (volume of mixing vessel[c])
        - sealed or open mixing vessel
        - volume of headspace
        - method of mixing, including rotor speed, etc.(mixing energy[c])

Minimum reporting requirements for ecotoxicology studies

  - duration of mixing and settling/separation time
  - method of phase separation in WAF preparation
  - type of chemical dispersant (if used)
- The exposure regime for the test substance (e.g., static, semi-static or flow-through) with details about the renewal regime and method.
  a. Define the exposure period[c]
  b. Approach used for establishing concentration dependent treatments: prepared at multiple loadings or prepared at single loading and then serially diluted[b]
  c. Distinguish between continuous or declining exposure[b]
- The frequency of exposure (e.g., how often the test substance is administered or renewed) and type of samples analyzed to determine toxicant concentrations.
- The method/design for determining the order in which test organisms are added to test vessels and the placement of test vessels (e.g., randomized, stratified random or Latin square design).

Number of replicates?

- The degree of replication of each treatment and control and an explanation of whether they are true replicates or pseudo-replicates.

Other

- Route of exposure to test organism should be clearly stated (e.g., via the diet, via the media, etc.).
- Details of all quality assurance and quality control procedures conducted as part of the study (e.g. whether the design was blind or double blind, whether scoring, data entry and calculations were conducted by one, or more than one individual, and see Section 4: Experimental Conditions as they relate to water quality, etc.).

3. Test organism characteristics

Name, source, and strain of species reported?

- Species selection – Justify the selected organism (e.g., ecological relevance and/or relevance to hypothesis).
- Identity of the species – Report the common and scientific name of the species, general type (e.g., plant), its source, and strain (if appropriate). Provide the DNA Bar-Code details if available. When dealing with new or cryptic species, genetic identification is recommended, e.g., the alga *Oophila sp.* (Baxter et al., 2015) or the *Hyalella azteca* species complex (Leung et al., 2016).
- Source – In-house cultures, wild populations and, where applicable, include method of collection (e.g., collection of fertilized eggs, or animals from the wild) and their subsequent handling.
- Life-history – The stage of the life-cycle, the age and sex of the test organisms, the number per treatment, and their size or mass at the beginning of the experiment, should be provided.

Control performance criteria met?

- Test species performance – Available historical data on endpoints (e.g., growth rates, reproduction) used in the experiment should be provided, enabling the data collected during the experiment to be put into context. For example, a detectable change in growth may not be deemed biologically significant when compared to historical control data of the performing laboratory (Länge et al., 2001). Control performance should be reported in order to permit comparison to validation criteria

Husbandry protocols listed?

- Husbandry – All procedures related to maintaining the organisms in good condition should be stated. Overviews of welfare and ethical approvals need to be reported, especially those that may influence observed responses (e.g., degree of enrichment, groupings). Outbreaks of disease or unexplained morality/morbidity, including their incidence and severity, and how these were treated, must be reported.

Minimum reporting requirements for ecotoxicology studies

4. Experimental conditions

General test conditions reported?

- General testing facilities – For example, growth chamber (make, model), tank (dimensions, capacity, rate of water change), mesocosm (location, volume, flow rate), greenhouse (location, size), field location (GPS coordinates, general climatic/environmental information for duration of study).
- Test conditions – All available details on relevant experimental parameters such as light intensity, photoperiod, and temperature should be reported as means, with the variability.
  a. Test vessels, e.g. open or closed, volume of headspace, addition of food[b]
  b. Inclusion of natural or simulated sunlight, i.e. wavelength, intensity, duration of ultra-violet light[b]

Source and condition of media?

- Source, type and composition of test media – e.g., water, commercially available media, soil, sediments, including known background contaminants.
- Test media parameters – All measurements that can influence test organism health or change endpoint responses should be reported (e.g., dissolved oxygen concentration, temperature, pH, salinity) as means with an estimate of variability (i.e., confidence intervals). In addition, report the properties of the test media that may influence interpretation of the test results (e.g., dissolved organic carbon where binding of test material is expected).
- Dosing mechanism – e.g., via the diet, peristaltic pump, spray application.

Acclimation and feeding?

- Details of acclimation – For both the test system and test organisms. In terms of the test system, this is to demonstrate that the conditions are stable prior to introduction of test organisms. This is of particular importance in mesocosm and sediment studies. In terms of the test organisms, this is to ensure survival and growth under experimental conditions (further details see Section 3: Test Organism Characteristics).
- Feeding – Information should include the type of food, source, amount provided and frequency of feeding. In the case of commercial foods, detailed reporting of characteristics is required. The concentration of any contaminants should be included, where relevant.

Other

- Number and density of organisms – This may be influenced by purpose of the test or experimental design (e.g., test power, see Section 8: Statistical Analysis). Nevertheless, test design should enable normal behavior of the test organism. Density of the test organisms will determine adequate feeding requirements and acceptable loading.
- Good Laboratory Practices (GLP) studies – When data are from a GLP study, this should be expressly stated in the publication, as well as the location of the raw data.
- Quality Assurance and Quality Control (QA/QC) activities – (e.g. calibration of laboratory equipment) and the results of these should be reported. If Standard Operating Procedures (SOPs) are used provide location of these.

5. Exposure confirmation

Clear statement of which samples were analyzed?

- Provide sufficient instrumental and details around your analytical approach, or a suitable reference, that supports the approach taken.
- State clearly what samples were analyzed (e.g., stocks only, exposure vessels, pooled or un-pooled) and the timing or frequency of measurements. This is important when interpreting the relevance of the observed response in light of actual exposure duration, regardless of test length.
- State the media type and the volumes sampled, as well as storage conditions and time till analysis.

Method LOD and LOQ provided?

- Report any relevant QA/QC undertaken during the sampling and analysis (e.g., blanks, storage studies, internal standards, recovery efficiency, and specific storage preservation techniques) and the results of the QA/QC.
- Report limits of detection, quantification, or reporting (LOD, LOQ, and LOR, respectively) and their variance.

Nominal or measured used in subsequent analyses?

- Report values in metric units as the target analyte, and not as the formulated product. Where applicable, provide means of measured values and

Minimum reporting requirements for ecotoxicology studies

standard deviations/errors.

Exposure metrics identified by Redman and Parkerton (2015): [b]

a. Oil loading
b. Total organic carbon (TOC)
c. Total petroleum hydrocarbon (TPH), e.g. IR, GC-MS, Fluorescence, check methods used)
d. Total petroleum aromatic hydrocarbons (TPAH)
e. Specific hydrocarbon constituents (BTEX, naphthalene)
f. Tissue residues of total or specific hydrocarbon classes (LC GC-MS)
g. Toxic units (TUs) based on measured or modeled dissolved hydrocarbons

Other

- State clearly whether subsequent statistical analyses and interpretations rely on nominal or measured values (See Section 8), and whether these exposures values have been corrected for recovery.
- Prepare a plan for archiving your original data (See Section 9).
- Size distribution of oil droplets[c]

---

6. Endpoints

All endpoints monitored, regardless of response, provided?

- State all endpoints monitored in the study, regardless of the observed response (e.g., avoid reporting only ‘differences’)
- Express clearly when and how the endpoint was monitored and recorded (e.g., blind evaluations of behaviour; See Section 2: Experimental Design) and how these data are presented in the paper (e.g., Tables, Supplemental information).

Clear definition and measurement units provided?

- Define the endpoint in order to remove ambiguity (e.g., what is a ‘malformation’?).

Other

- Justify the selection of your endpoints (also see Section 2: Experimental Design) and their statistical power (see Section 8: Statistical Analysis).
- Report other observations that may have relevance, but were not an explicit part of the original study design (e.g., lesions in fish), as this can inform future work and be hypothesis generating.

---

7. Presentation of results and data

All data, regardless of statistical significance is discussed?

- Inclusion of statistically and biologically significant, as well as non-significant results, will allow for a balanced understanding of the full range of responses.

Untransformed data provided?

- It is useful to report summary data (e.g., end-points, estimates, ranges) in table-format, and greater attempts should be made to include as much data as practical, e.g., in Supplemental Information (See Section 9: Raw Data).

Other

- Create figures that provide readers with greater ability to assess data distributions and variability (e.g., scatterplots, histograms, box plots) for each time-point. It is common for researchers to rely upon graphs displaying a mean ± standard deviation or standard error, but this is problematic because different distributions of data can be represented in the exact same way when relying solely upon bar charts and line charts (Weissgerber et al., 2015). Furthermore, traditional bar charts can disguise outliers in data, which can be important, particularly for studies with small sample sizes (Weissgerber et al., 2015).
- Employ appropriate scales (e.g., do not truncate or break axes to over-emphasize effect sizes or differences relative to controls).
- Provide details in figure or captions specifying the statistical test employed (if applicable), degree of replication, level of statistical significance (if any).

Minimum reporting requirements for ecotoxicology studies

- Confidence intervals, with alpha-level, should accompany summary statistics in figures and tables, instead of standard deviation and standard error. Confidence intervals are preferred over standard deviations or standard errors because measures of uncertainty are not always symmetric about the mean. This is especially relevant when the data are transformed for analysis and the results are expressed in back-transformed values. It also applies when some non-normal error structures are used, such as Poisson or binomial. Only in the case of normally distributed data, will standard deviations provide equivalent information to confidence intervals.
- For field studies it is helpful to provide maps with relevant information (e.g., GPS coordinates, scale bars, orientation, regional context) that may enable a reader to understand the spatial context of a study.

8. Statistical analysis

Statistical flowchart?

- Provide a statistical flow chart that includes any preliminary data checks to satisfy test or model requirements and accommodations (e.g., transformations or robust methods) of data problems, the handling of multiple controls (e.g., solvent and negative controls), and indicate how statistical tests or models are selected and why (e.g., OECD, 2006, 2012).

Transformations justified?

- Provide justification for any transformations (e.g., logarithm) used and whether/how it affects the analysis results.

All outliers are reported

- Statistical outliers should be identified and their effect on conclusions should be stated.

Justification for model selection and variables?

- For complex models (for regression or hypothesis testing) with multiple potential explanatory variables, any model selection method that sequentially adds or removes terms to arrive at a final model should be described and, if possible, verified by alternative approaches to avoid unintentional bias.
- Model selection and goodness-of-fit criteria should be specified.

NOE-LOEC: power of test and percent change reported?

- Power to detect a specified size effect should be stated.
- Confidence intervals for the mean response at the NOEC and LOEC should be reported along with the percent change from control.
- If the response is non-monotonic (e.g., hormesis), indicate how this was addressed in the statistical analysis
- Define percent change from control (the ‘x’ in ECx) as to whether it applies to raw or transformed data.

ECx: Model estimates and confidence intervals provided?

- Report when ECx estimates are extrapolations, including extrapolations below the lowest tested positive concentration/dose.
- Confidence intervals should be reported for ECx and all model parameters and explain why any parameters not significantly different from zero were included in the model, as these can indicate possible model problems.

Other

- Power of the planned hypothesis testing procedure to find effects or the ability of a regression model to estimate ECx reliably should be re- ported (See Section 4: Experimental Conditions). This can be done partly through the use of historical control data (See Section 3: Test Organism Characteristics).
- Good estimation of the control mean is important since all tests and estimates are in relation to that mean, so if there were more replicates in the control than in treatment groups or other special considerations of the control, make clear what was done and why.
- In reporting sub-lethal effects in a study with substantial mortality in high treatment groups or loss of subjects/replicates for other reasons, report any adjustments made to the tests or models (e.g., weighting) to avoid over-interpretation resulting from small sample size.
- Explain how the statistics account for the actual experimental set-up (e.g., individual, paired or group housing, expected monotone dose-response or deviations therefrom, such as hormesis).
- Results should be reported in the original units and with no more significant digits than the raw data justify. For example, if the data are measured with two significant digits to the right of the decimal point, it is pointless to report means to five decimals points. That implies a level of precision not justified by the data. Moreover, the quality of the measurements determines how many significant digits are meaningful. If the equipment being used to measure

Minimum reporting requirements for ecotoxicology studies
a response is ac- curate only to the nearest 0.1, but reports five digits past the decimal point, data should be reported only to the nearest 0.1. Summary statistics should be reported as means and confidence intervals (not standard deviations or standard errors) with an explanation of how confidence intervals were determined. This is especially important if transformed data were analyzed, so confidence intervals are not symmetric about the mean. • While expressing change from control (percent of control) is useful for presentation, data should also be presented as actual recorded values. • Report what exposure values (e.g., measured versus nominal) were used in your estimates. • The minimum effect size that a regression model can reliably estimate should be determined and reported. For example, if the standard error of the control mean is 20% of that mean, then the estimation of ECx for x b 20 is likely unjustified.
9. Raw data Nominal and measured concentrations provided? • The replicate and treatment identification. • The measured and nominal chemical concentrations for each analysis. Untransformed response by replicate available in some form? • The non-transformed (e.g., non-logged, non-normalised) biological effects data at the level of the unit of measurement. For example, data on individuals if that is what is measured or data for groups of individuals when pooled and then measured. *Other* • All measured values for experimental conditions that are known to affect the toxicity or bioavailability of the chemical.