EVALUATION AND DEVELOPMENT OF AQUATIC TOXICITY TEST METHODOLOGIES

by

Amy Beasley

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy in Molecular Bioscience

Middle Tennessee State University
May 2015

Dissertation Committee:

Dr. Ryan R. Otter, Chair

Dr. Scott E. Belanger

Dr. Frank C. Bailey

Dr. Eric Van Genderen

Dr. Jason R. Jessen

This dissertation is dedicated to my children: Savannah Lee, Tyler, Will, and Megan. You four are smart, fun, kind, creative and the best team ever. I wanted to do this well for you -- to set an example of hard work and determination and also to provide for your futures. This work is a great accomplishment, but what we've built together as a family is hands down the greatest. Love you.

Also to my mother Lou who told me with uncharacteristic bluntness, "If you don't take this opportunity you're an idiot." You were right. Thank you for daily support and encouragement, tangible and otherwise. I'm grateful beyond words.

ACKNOWLEDGMENTS

Thanks to the Middle Tennessee State University Biology Department and the Molecular Biosciences Program. Thanks also to the MTSU Undergraduate Research Center for the Undergraduate Research Experience and Creative Activity (URECA) grant which funded my first research endeavor.

I'm deeply grateful to my incredible, diverse committee for their support and unique perspectives: Dr. Eric van Genderen, Dr. Frank Bailey, and Dr. Jason Jessen. Dr. Scott Belanger has been my faithful advocate and a priceless professional mentor; I am thankful that we will continue our work together.

I chose Dr. Ryan Otter as my graduate advisor because I knew he would be satisfied with no less than my best and would allow me to be creative and innovative in my work. He has also given me no less than his best and has mentored me with great integrity. He has my greatest respect and gratitude.

ABSTRACT

International chemical management legislation such as the European Union's REACH (Registration, Evaluation, Authorisation, and Restriction of Chemicals) has increased the need for more and better toxicity data while recognizing the 3R's of animal use (reduction, replacement, and refinement). To address this need, the Organisation for Economic Cooperation and Development's (OECD) Fish Toxicity Framework guidance document recommended improvements to test guidelines and integrated testing strategies. This study contributes to recommendations outlined in the Fish Toxicity Framework and similar guidelines. Chapter I introduces the Embryonic Developmental Rate Assay (EDRA), a streamlined approach that uses nonlethal endpoints to assess toxicity. Time-lapse video technology was used to track the timing of seven developmental stages covering embryogenesis through hatching. Embryos exposed to two known toxicants demonstrated that developmental timing and rate were effective biomarkers for toxicity.

Another refinement to test methodology is presented in Chapter II. The Stepwise Information-Filtering Tool (SIFT) is a systematic method to break down large toxicity datasets for analysis in a stepwise manner**,** applying user-defined criteria to address reliability and relevance. A case study application of SIFT to analysis of a chronic daphnid toxicity dataset was presented, as well as a comparison to two similar data quality methodologies.

In Chapter III, the relationships between common summary statistics in toxicity testing were evaluated. The widely used hypothesis-based NOEC (No Observed Effect Concentration) has critical issues (e.g., impact of test design, no confidence levels). The ECx (Effect Concentration, with x as the percentage effect compared to controls) is regression-based, where x at a low percentage (e.g.10 or 20%) is considered an analogue to the NOEC. Using a chronic daphnid toxicity dataset, the relationship of NOEC to EC_{10} and EC_{20} and the impact of test parameters to the relationship was evaluated.

Threshold of toxicological concern (TTC) concepts have been used for years to estimate 'safe' concentrations for low-volume chemicals or additives, although this concept has not yet translated to ecotoxicological thresholds (ecoTTC). In Chapter IV, a comprehensive dataset was constructed, analyzed for ecoTTC values, and evaluated for the influence of chemical class, endpoint selection, and regional application factor choice.

TABLE OF CONTENTS

	Page
LIST OF TABLES	viii
LIST OF FIGURES	x
INTRODUCTION	1
Hazard/Risk Assessment	1
Regulatory Environment	4
Animal Alternative Testing	6
OBJECTIVES	8
CHAPTER I: Embryonic Developmental Rate Assay (EDRA) for screening and evaluation of developmental timing in zebrafish as a biomarker of toxicity	9
Introduction	9
Methods	12
Results	23
Discussion	26
References	31
CHAPTER II: Stepwise Information-Filtering Tool (SIFT): A method for using risk assessment metadata in a nontraditional way	35
Introduction	35
Methods	37
Results	44
Discussion	47

Conclusion 51

References 52

CHAPTER III: Evaluation and comparison of the relationship between NOEC and EC_{10}/EC_{20} values in chronic *Daphnia* toxicity testing 55

Introduction 55

Methods 58

Results 63

Discussion 68

Conclusion 74

References 76

Appendices 80

CHAPTER IV: Influence of chemical class, endpoint selection and application factor parameters on thresholds of toxic concern in ecotoxicological hazard assessment 85

Introduction 85

Methods 87

Results 88

Discussion 93

Conclusion 95

References 97

PROJECT CONCLUSION 99

References 101

LIST OF TABLES

Page

CHAPTER I

Table 1. Time in minutes to reach each of the 7 previously identified morphological stages. 24

CHAPTER II

Table 1. SIFT criteria developed specific to the purpose of the endpoint comparison case study. 40

Table 2. Comparison of Agerstrand versus SIFT criteria as applied to the case study dataset. 41

Table 3. Criteria developed by Klimisch et al. (1997) to evaluate reliability. 43

CHAPTER III

Table 1. Information collected from 210 Daphnia chronic toxicity test reports and peer-reviewed studies to populate a master dataset. 61

Table 2. Analysis of NOEC:EC_{10} and NOEC:EC_{20} relationships for the Test Type parameter. 65

Table 3. Analysis of NOEC:EC_{10} and NOEC:EC_{20} relationships for the Species parameter. 65

Table 4. Analysis of NOEC:EC_{10} and NOEC:EC_{20} relationships for the Strategy parameter. 66

Table 5. Analysis of NOEC:EC_{10} and NOEC:EC_{20} relationships for the pH parameter. 67

Table 6. Analysis of NOEC:EC_{10} and NOEC:EC_{20} relationships for the Hardness parameter. 67

Table 7. Analysis of NOEC:EC_{10} and NOEC:EC_{20} relationships for the Temperature parameter. 68

LIST OF FIGURES

Page

INTRODUCTION

Figure 1. Process diagram of hazard and risk assessment to eventual risk management. 2

Figure 2. Sample aquatic hazard assessment methodology. 4

CHAPTER I

Figure 1. 64-cell stage with dividing cells massing on yolk sac. 15

Figure 2. Sphere stage. 16

Figure 3. Shield stage. 17

Figure 4. Bud stage. 18

Figure 5. Eye/7-somite stage. 19

Figure 6. Movement stage. 20

Figure 7. Otolith stage. 21

Figure 8. Hatching stage. 22

Figure 9. Measurement of shortest distance from eye to otic vesicle in microns. 23

Figure 10. Mean time to reach morphological stages. 25

Figure 11. Ear-eye distance measurement. 26

CHAPTER II

Figure 1. Process diagram of the SIFT methodology. 39

CHAPTER III

Figure 1. Most sensitive endpoints for calculated effect concentrations. 64

CHAPTER IV

Figure 1. Process diagram showing the progression of ecoTTC construction and refinement. 89

Figure 2. Distribution of a general subset of the master PNEC dataset. 91

Figure 3. Distribution of a subset of cationically charged chemicals from the master PNEC dataset. 92

Figure 4. Distribution of a subset of cationically charged chemicals from the master PNEC dataset. 93

INTRODUCTION

HAZARD/RISK ASSESSMENT

Risk assessment is the common term describing the entire process of chemical hazard and risk identification, characterization, and evaluation. Hazard refers to the potential danger inherent in response to a chemical, while risk refers to the likelihood of exposure to that hazard. Assessments are commonly performed in industry and government for both new and existing chemicals. The purpose is to not only understand hazard and risk but to also accurately account for the inherent uncertainty that exists in any assessment. Hazard and risk assessments follow similar but distinct pathways before coming together in a final risk characterization step (Figure 1).

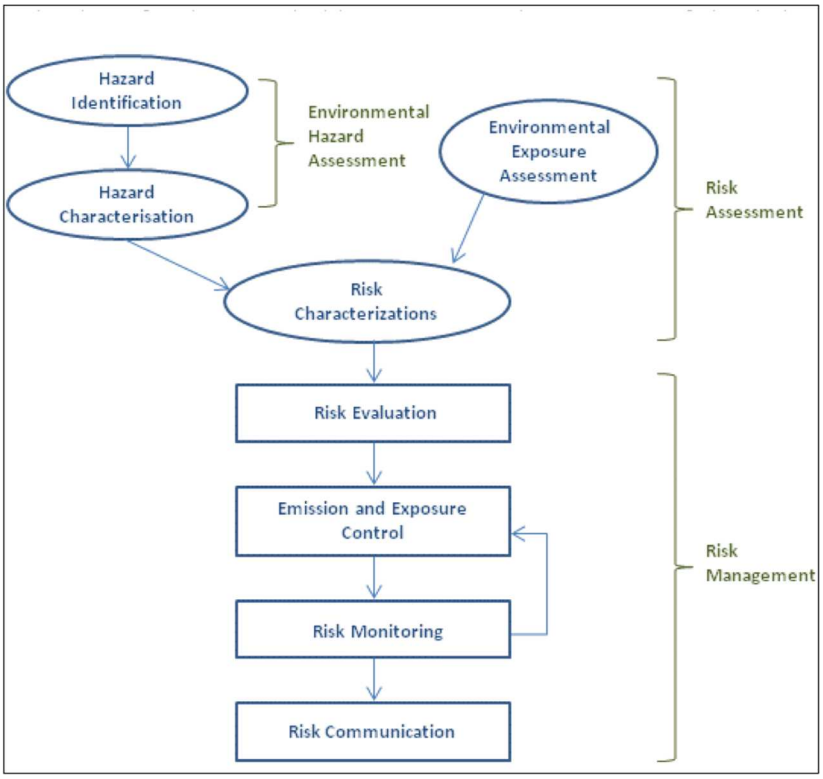

Figure 1. Process diagram of hazard and risk assessment to eventual risk management. Adapted from OECD Toolbox, 2014.

Hazard identification and characterization begins with collection of available toxicology data. Data are evaluated individually for reliability and relevance, then the collection is evaluated for adequacy and any gaps identified. Endpoints are identified that can represent potential hazard in the appropriate compartment (e.g., aquatic, terrestrial, marine, or sediment effects). Additional data can be generated using appropriate standardized toxicity testing methods as well as predictive modeling (e.g., QSAR quantitative structure-activity relationship models).

In an aquatic hazard assessment, the minimum relevant data ideally includes information from representative species at three trophic levels: acute toxicity to fish, acute toxicity to invertebrates (*Daphnia*), and acute toxicity to algae. QSAR or other model data is incorporated to calculate a Predicted No Effect Concentration (PNEC), which is then used as either a threshold value within a hazard assessment or as part of the benchmark value in a risk assessment. The PNEC represents a concentration at which the test chemical is 'safe' for the exposed population. This PNEC is refined with additional acute toxicity data or data from chronic exposures (Figure 2). Application factors based on regional guidances are used to reflect the appropriate level of uncertainty from sources such as species-to-species or lab-to-field extrapolation [3]. As the data collection is expanded and refined, the application factor decreases (Figure 2). The PNEC is combined with a Predicted Environmental Concentration (PEC) similarly generated from the risk assessment methodology to give a ratio predictive of the risk to exposed populations.

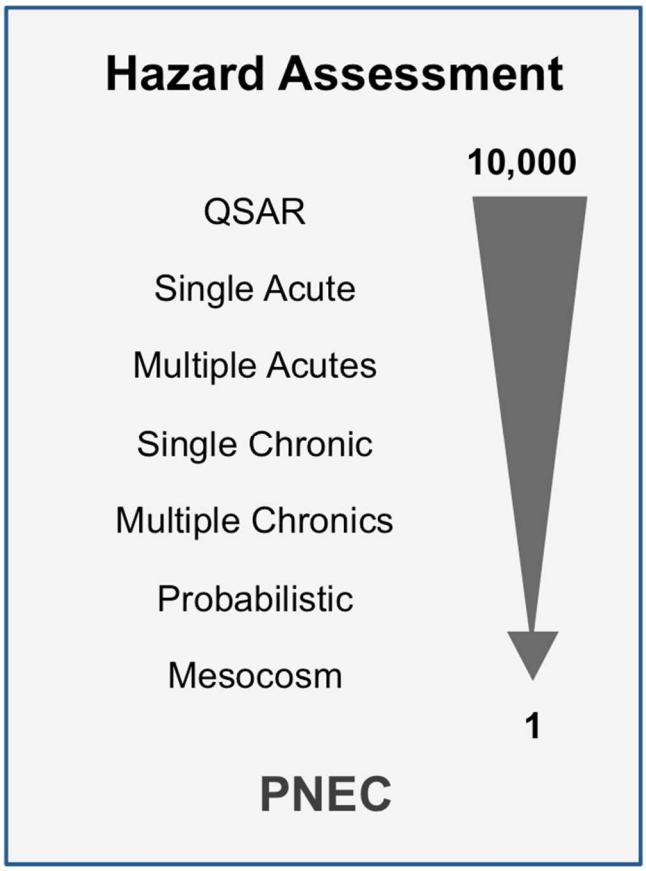

Figure 2. Sample aquatic hazard assessment methodology. Potential application factors decrease from a maximum of 10,000 as data collection is refined for extrapolation to predict effects in the environment.

REGULATORY ENVIRONMENT

A comparison study conducted by the OECD in 2004 highlighted similarities in risk assessment approaches among member nations including the US and UK; both conducted hazard assessments paired with a separate (usually quantitative) exposure assessment [4]. Differences in governance and policy exist between governments within the general approach to risk assessment.

US chemical review and risk assessment is based on the Toxic Substances Control Act (TSCA) [5] and administered by OCSPP (Office of Chemical Safety and Pollution Prevention) of the United States Environmental Protection Agency (USEPA). Acute toxicity data is the foundation of a tiered evaluation system, where potential risk determines placement in subsequent tiers requiring additional toxicity data. Standardized testing guidelines developed by USEPA are recommended for generation of toxicity data for application within the United States. TSCA requires that new chemicals, whether developed or imported, must be reported via premanufacture notice. USEPA then screens for potential risk and delivers a clearance or restriction based on those findings. Existing chemicals are also regulated based on TSCA, with the addition of updated chemical management programs to increase transparency, data availability and regulation of high-risk target chemicals.

The European Union uses a top-down approach to chemical review and risk assessment, where manufacturing or importing tonnage determines the level of toxicity information and testing required. EU member nations enacted legislation in 2006 to restructure and improve chemical management [1]. Under REACH (Registration, Evaluation, Authorization and Restriction of Chemicals), industry takes responsibility for toxicity testing and registration of chemicals with the European Chemicals Agency (ECHA) based on tonnage. ECHA evaluates the submitted chemical dossiers, ranks them based on potential effect, then

authorizes or restricts use. Overlaying the REACH framework, the OECD (Organization for Economic Cooperation and Development) provides data and tools to help countries institute policies for environmental risk management. Test guidelines are standardized and harmonized for international use and to allow mutual data acceptance among the 34 OECD member countries.

ANIMAL ALTERNATIVE TESTING

The 3 R's (reduction, replacement, refinement) are increasingly shaping the landscape of global chemical risk assessment [6]. In environmental toxicity testing, fish are a key component and the focus of building concern over animal use. OECD guidelines include the newly adopted Fish Embryo Test for acute toxicity [7], the Fish Early-Life Stage Test for chronic toxicity [8] and the Fish Short-Term Reproduction Assay [9]. Increased chemical regulation under REACH has increased the use of fish in EU member states although overall REACH has reduced the number of animals used in toxicological testing since 2008 [10]. Effluent testing with fish is restricted in many European countries (e.g., Germany) or has been eliminated based on ethical grounds (e.g., United Kingdom). However, effluent testing is still the largest source of fish use, accounting for 3-6 million fish annually [11].

In 2012, the OECD released the Fish Toxicity Framework review of fish toxicity testing processes, at the same time addressing the need for further reductions in whole-fish testing [2]. In it, Integrated Testing Strategies are highlighted as a

'more efficient framework', incorporating all available information including modeling, *in silico, in vitro*, and read-across methods. Also highlighted in the Framework is the need to optimize existing toxicity testing strategies to reduce fish use while retaining the test power and therefore usefulness of the test. Nonlethal endpoints (e.g., moribund vs. death), species sensitivity distributions (e.g., the USEPA Interspecies Correlation Estimation tool), and threshold approaches are discussed as potential sources of reduction, replacement and refinement to the Fish Toxicity Framework.

OBJECTIVES

1) To create a simple, straightforward method to assess toxicity in zebrafish using developmental rate as a biomarker, and to demonstrate the utility of the method with known toxicants.

2) To develop a systematic method to sift through and select the best, most reliable and relevant data from large toxicity datasets in order to probe questions other than the typical risk assessment, and to apply this method in a case study of chronic toxicity test data.

3) To conduct a data-driven evaluation of summary statistics (ECx and NOEC) crucial to aquatic toxicity testing, including impact of test parameters and design, and recommend changes to standardized test guidelines based on the findings.

4) To construct a metadataset, calculate preliminary ecoTTC values from that metadata then perform a subsequent analysis of the impact of endpoint, application factor, and chemical class on calculated ecoTTCs.

CHAPTER I

Embryonic Developmental Rate Assay (EDRA) for screening and evaluation of developmental timing in zebrafish as a biomarker of toxicity

INTRODUCTION

Zebrafish (*Danio rerio*) are a commonly used model organism in toxicity testing. Although the fathead minnow (*Pimephales promelas*) has historically been the North American model organism of choice, *Danio* has in recent years gained popularity in the U.S. and is the favored model organism in Europe [1]. The Fish Embryo Test (FET) which was developed for use with zebrafish embryos was recently adopted by the Organisation for Economic Cooperation and Development (OECD) as a standard alternative to the traditional fish acute toxicity test [2]. Zebrafish have gained favor for many reasons including ease and cost of culture. Embryogenesis is achieved by 72 hours and is observable due to the transparent chorion that encases the embryo [3]. Further, development from 72-140 hours post fertilization (hpf) is considered within the eleutheroembryonic stage, which precludes European Union protections that begin with exogenous feeding [4, 5]. Females often produce 200 or more embryos per clutch and can be bred frequently. The zebrafish genome is annotated, reflecting 70% synteny with the human genome [6]. A variety of transgenic *Danio* strains are available for exploration of specific modes of action (e.g., endocrine disruption)[7].

Tests such as the OECD Fish Embryo Test and Fish Acute Toxicity Test use morphological indicators to determine mortality and therefore acute toxicity [2, 8]. Acute toxicity is based on the exposure concentration at which 50% lethality is observed (Lethal Concentration, or LC50). It is possible (even likely) that toxic effects are still occurring at sublethal concentrations below the LC50 [9]. These effects would be difficult to observe using the mortality endpoints from the FET/Acute tests (e.g. congealment, heart rate), but such effects could be very useful in determining an overall toxicity profile. In addition, the acute test timeframe (< 72h post fertilization, or hpf) provides a unique window into toxic mode of action and teratogenicity [10].

Developmental rate has been used as a biomarker to assess sublethal effects during embryogenesis [9, 11-13]. Tracking time to achieve morphological stages provides a quantitative measurement of changes to developmental timing. The addition of successive measurements of the distance between the eye and developing ear (EED) has been shown to provide useful quantitative information on developmental timing [3, 11, 14]. Onset of gene expression has also been used as a simple method to track changes to developmental rate [15]. New methods utilizing sophisticated microscopy techniques and automated image capture can reveal detailed morphological staging, resulting in more accurate descriptions of changes to developmental timing [16, 17]. However, such assays

require advanced equipment and time, therefore are more suited to exploration of particular morphological effects. In this study, a simple assay to screen and evaluate developmental timing changes for effects in embryonic zebrafish was developed. The Embryonic Developmental Rate Assay (EDRA) combines seven easily identified morphological stages with measurement of EED to provide a basic evaluation of developmental rate over the embryogenetic timeframe. Exposure to copper and tetradecyl sulfate, contaminants known to cause developmental delays in zebrafish [18], demonstrated the usefulness of EDRA.

Copper toxicity in early-life stages of teleosts, including zebrafish, is well characterized [10, 19-21]. The window of most profound toxicity appears to precede chorion hardening at 1-4 hpf and resume at hatching, when larvae are then directly exposed to contamination [19]. Despite the protective barrier of the chorion during hours 4 to hatch, effects of copper exposure occur. Sublethal effects to developing embryos include increases in heart rate, yolk sac size, and decreased length; these effects are suggestive of delays to overall development [21]. Delayed and inhibited hatching has been shown to occur in a dose-dependent manner with embryonic exposure to copper [21, 22]. Although hatching is a complex process, copper inhibits the function of the chorionase enzyme, which is crucial to chorion rupture [23].

Tetradecyl sulfate (TDS) is used as a well-known pharmaceutical treatment for varicose veins [24] and is also known as C14AS (alkyl sulfate), a common detergent range surfactant. TDS has been tested extensively in aquatic systems [25, 26], including the validation study of the FET [27] . FET validation testing showed that TDS was significantly more toxic in zebrafish embryos as compared to adult fish [27]. Developmental effects of exposure to TDS have been shown in fathead minnow, including impaired hatching [28, 29].

The objectives of this study were to 1) create the Embryonic Developmental Rate Assay using time-lapse image capture of zebrafish embryo early-life stages; and 2) evaluate the utility of EDRA with zebrafish embryos exposed to copper and TDS.

METHODS

Zebrafish culture

Danio rerio were cultured in 3.5L tanks of system water. System water was type 3 reverse osmosis with added salts to maintain 850ms conductivity. pH was maintained at 7.3 using sodium bicarbonate. Water temperature was maintained at 28° C ± 0.5 and light on a 14:10-h light:dark cycle. Fish were fed *Artemia* brine shrimp.

Embryo collection

Fish were bred twice weekly in mating tanks with perforated floors. Embryos were collected, rinsed with system water, and allowed to mature for 1 hour to assess fertilization and age. Embryos reserved for negative controls (untreated) were placed in groups of 15 in 10-cm Petri dishes and incubated at 28.5° C ± 0.5 and 14:10 light cycles.

Test chemical exposures

Test concentrations of tetradecyl sulfate were prepared according to methods found in the Annexes to Test Validation of the Zebrafish Embryo Toxicity Test [30]. Five embryos were treated with 15 mL of 0.3125mg/L tetradecyl sulfate for filming, and an additional 10 embryos were treated with 20 mL of 0.3125 mg/L tetradecyl sulfate as a positive control.

Test concentrations of copper(II) sulfate pentahydrate ($CuSO_4$ - $5H_2O$) were prepared according to methods in Annexes (OECD 2012 etc). Five embryos were treated with 15 mL of 0.30 mg/L copper(II) sulfate pentahydrate for filming, and an additional 10 embryos were treated with 20 mL of 0.30 mg/L copper(II) sulfate pentahydrate as a positive control.

Film staging

Staging dishes were prepared using 10cm Petri dishes with 2mm hemispherical wells cast in agarose gel. Image sequences were captured at 28.5° C and 14:10 light cycles. Each sequence was filmed using only one embryo aged <2 h post fertilization (hpf), using an Infinity-1 camera (Lumenera Corporation, Ontario, Canada) mounted to an inverted microscope (Nikon, Eclipse TS100, Japan) housed in an environmental chamber (Sheldon Manufacturing, Model 2015, Cornelius, OR). Frame capture rates were 1 frame/min for durations of 36-72 hours, depending on hatch time. Image sequences were analyzed with NIS-Elements software (Melville, NY, USA).

Developmental staging analysis

Seven morphological milestones were chosen to represent developmental stages between 2 h post fertilization and hatching (Figures 1-8). Time-stamped images were identified for each milestone and overall time (in hpf) to reach each milestone determined. Time in hpf to reach milestones determined overall developmental rate. Ear-to-eye distance (EED) was calculated by measuring the shortest distance between the eye and otic vesicle from 10 to 15 randomly selected frames from appearance of the otic vesicle to the end of the film sequence (Figure 9).

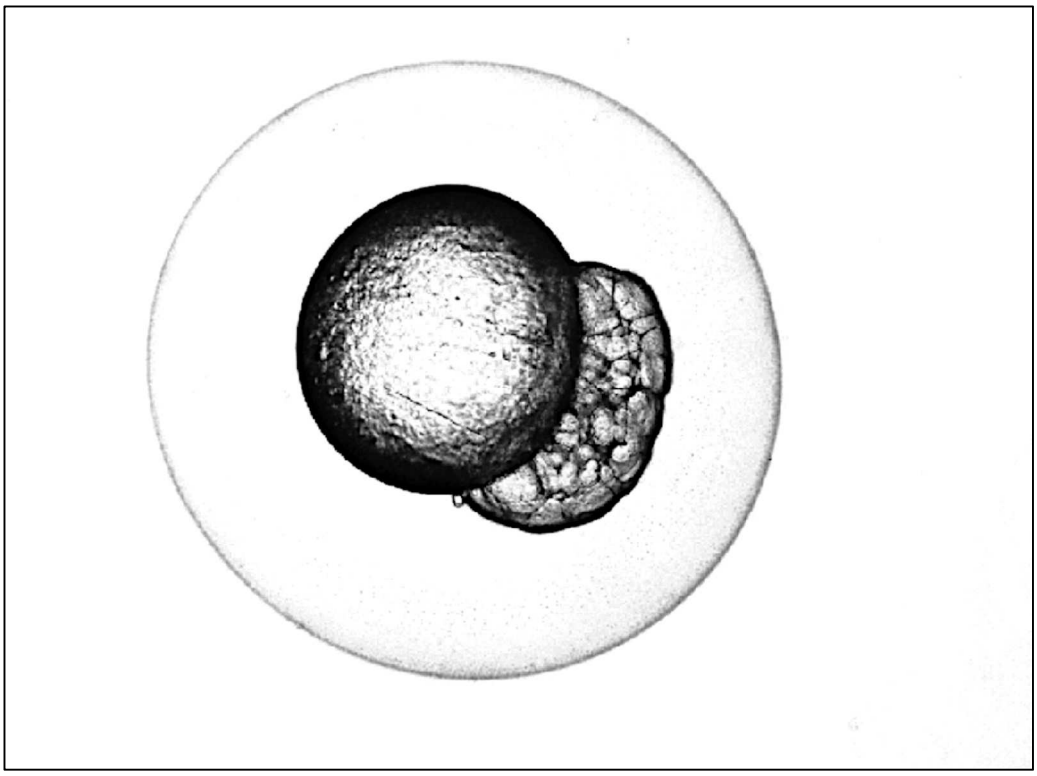

Figure 1. 64-cell stage with dividing cells massing on yolk sac.

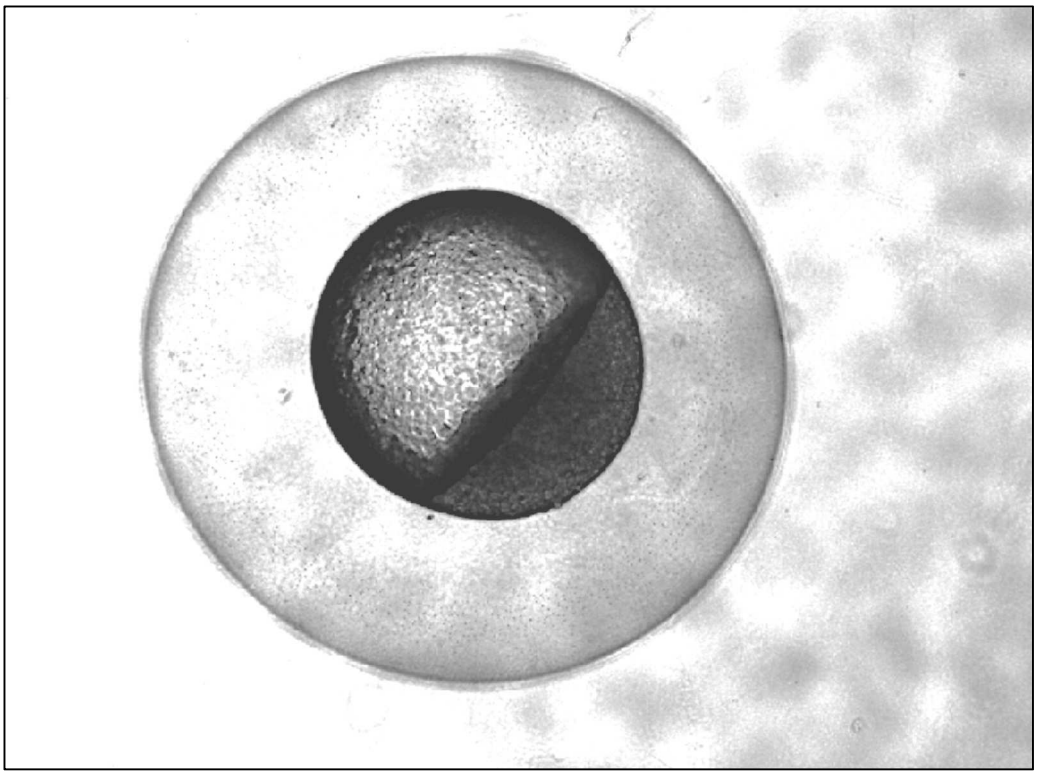

Figure 2. Sphere stage. Mass of dividing cells spreads and flattens at interface with yolk sac, creating a spherical appearance.

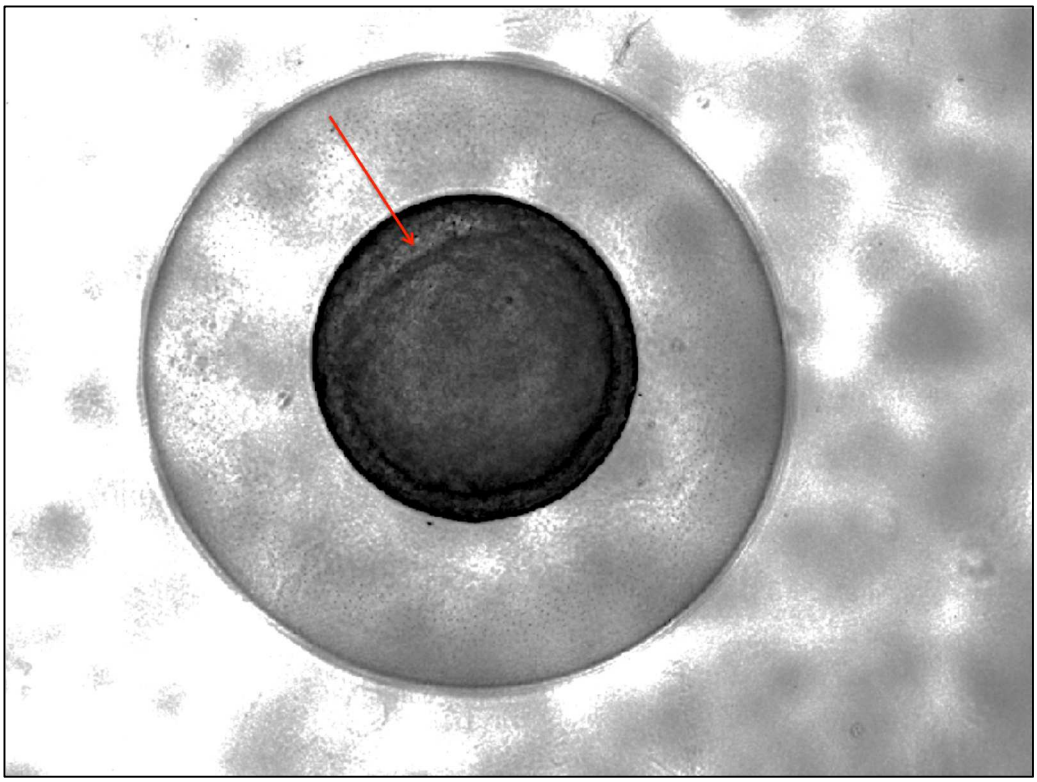

Figure 3. Shield stage. Cell mass cap covers 50% of the yolk sac and embryonic shield (noted by arrow) is visible.

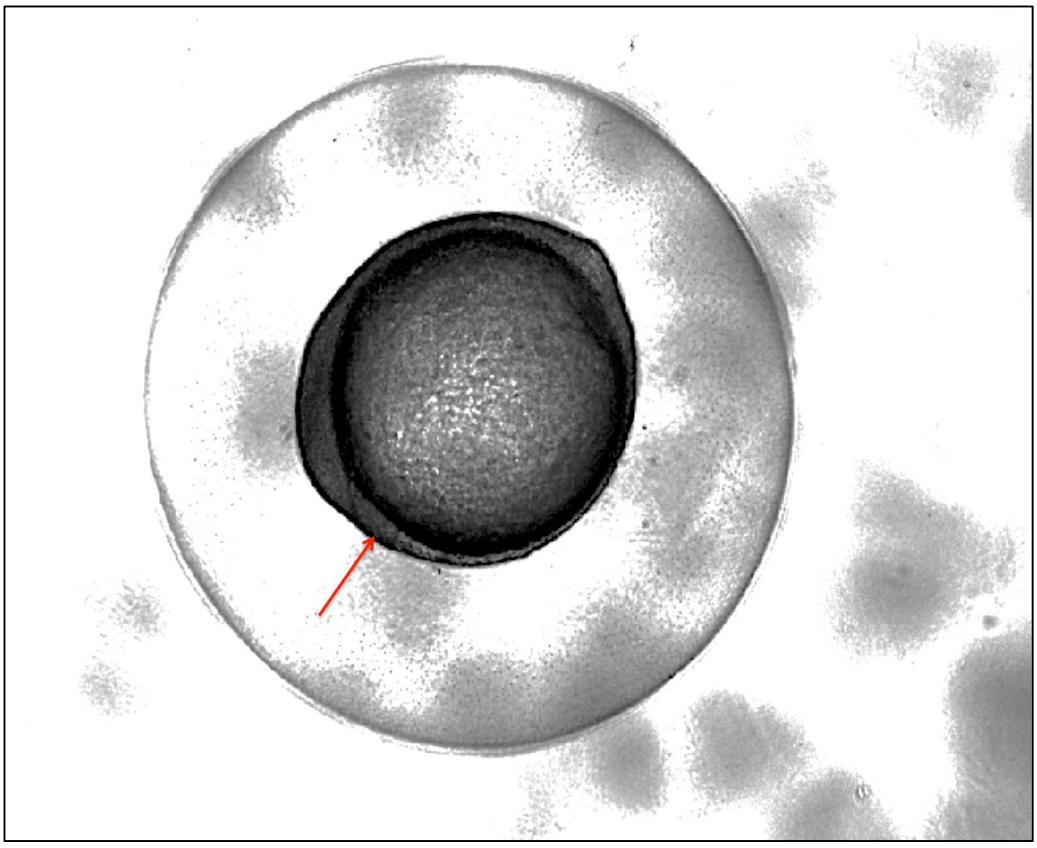

Figure 4. Bud stage. Tail bud and polster (precursor to the hatching gland, noted by arrow) are visible.

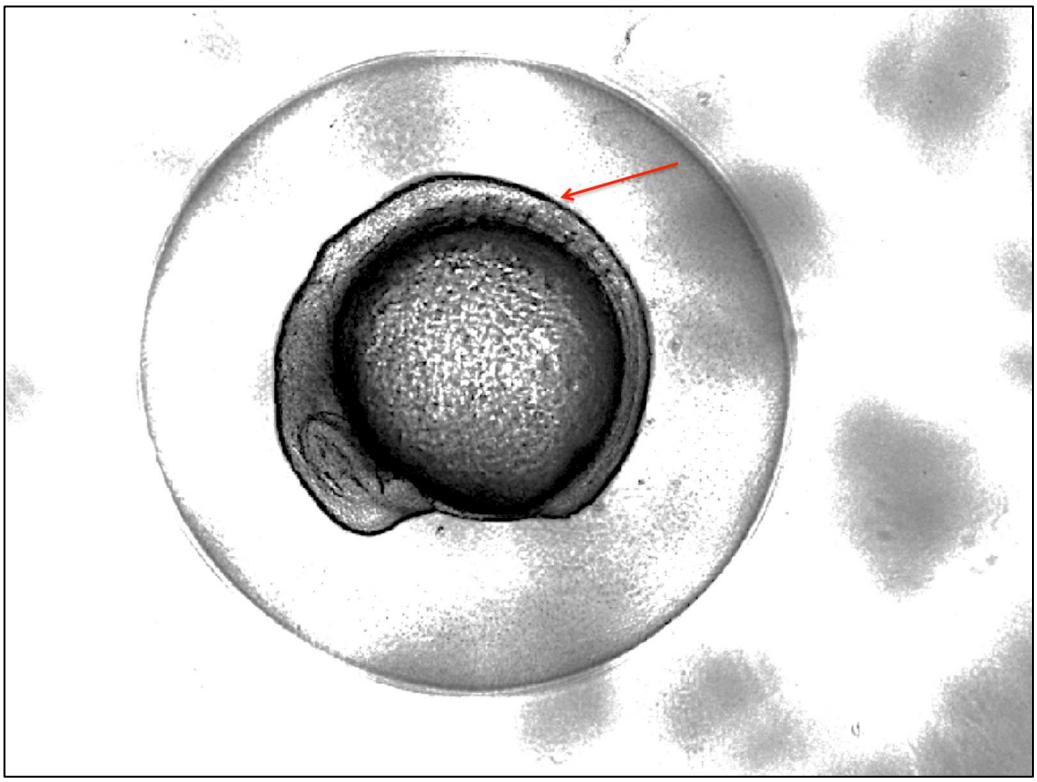

Figure 5. Eye/7 somite stage. Eye structures are visible. Seven somites (noted by arrow) are formed and visible.

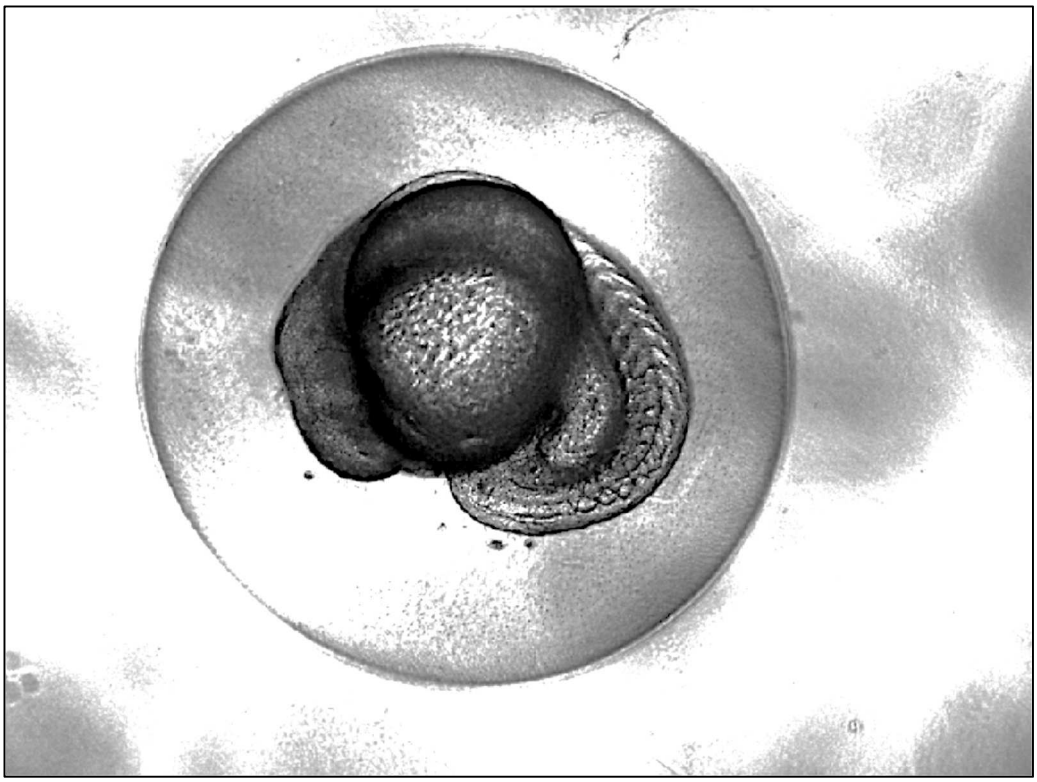

Figure 6. Movement stage. Tail separated from yolk sac and muscle twitches cause observable movement of the tail.

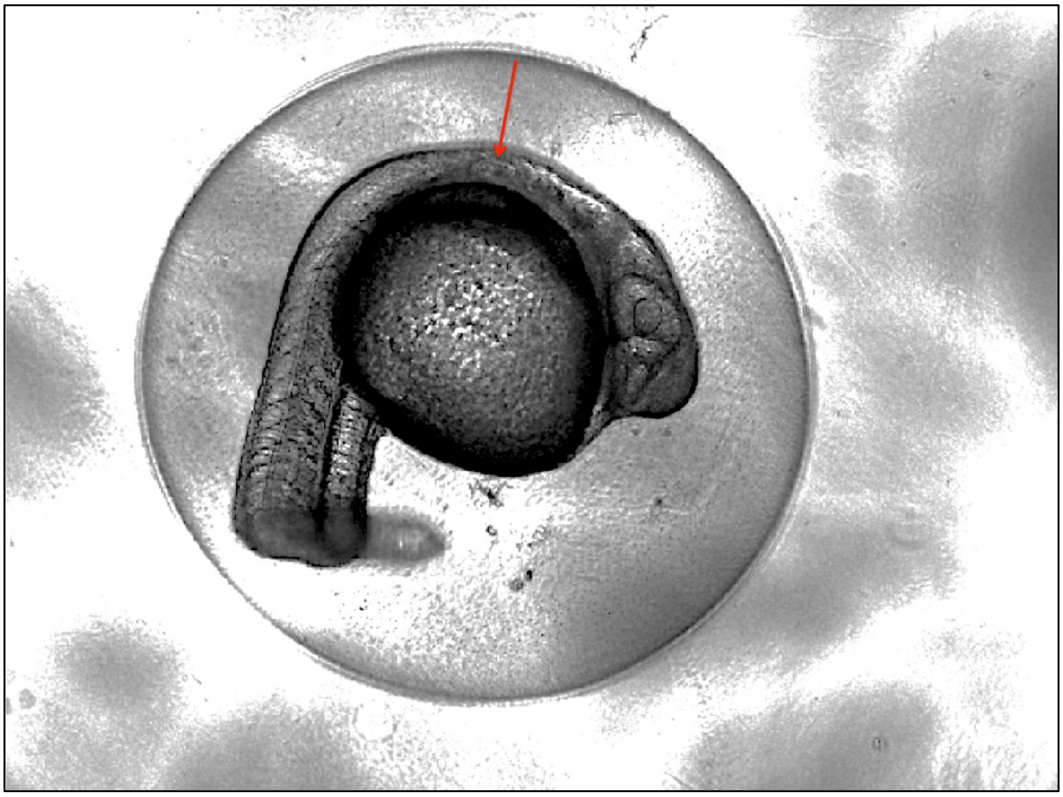

Figure 7. Otolith stage. Otoliths (noted by arrow) are clearly visible from the side.

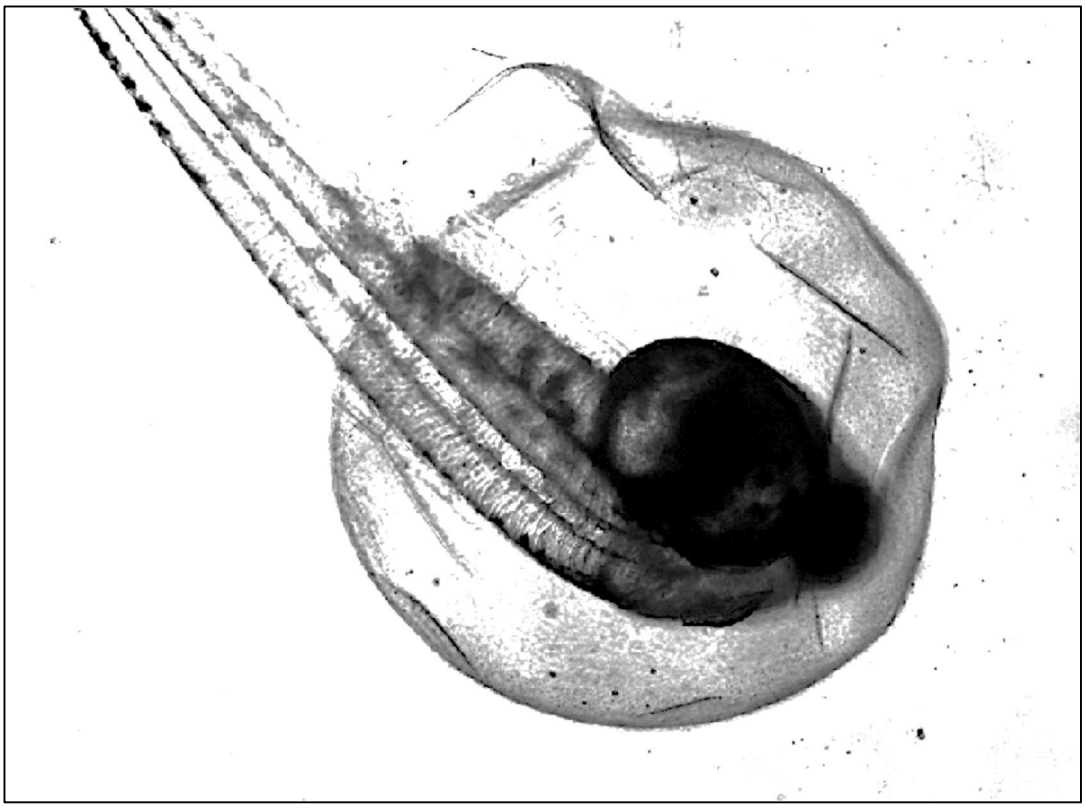

Figure 8. Hatching stage. Embryo has ruptured the chorion.

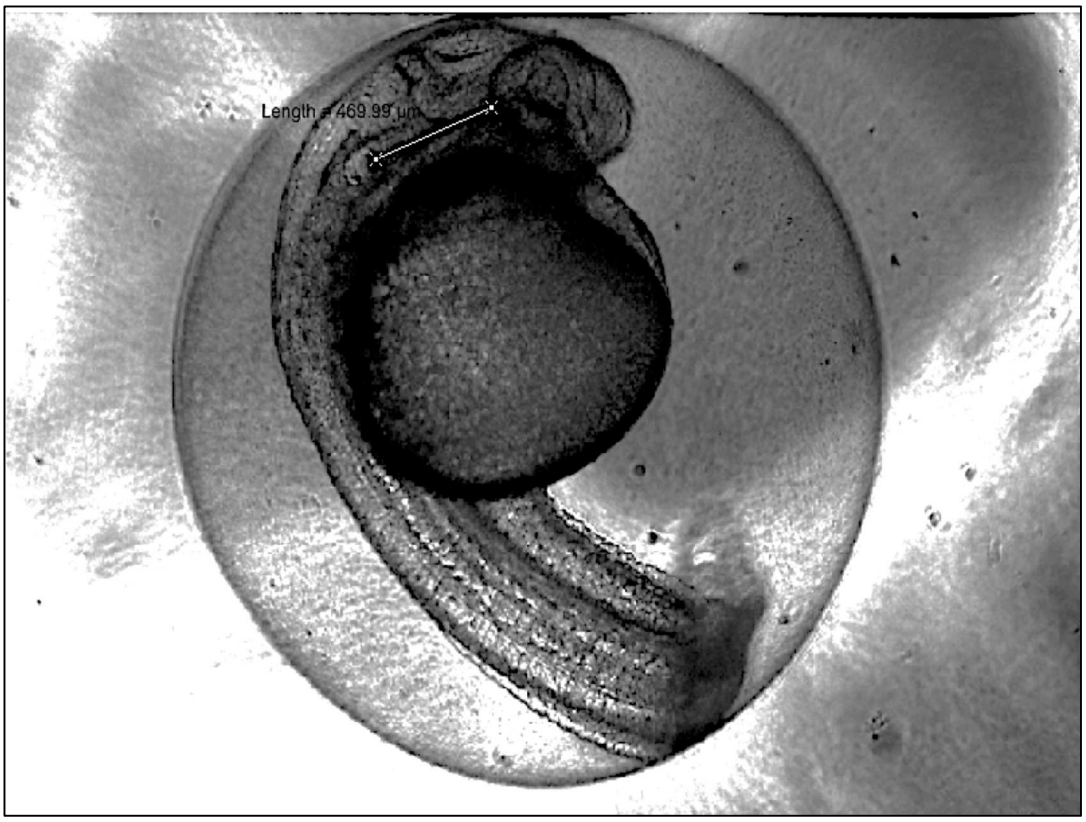

Figure 9. Measurement of shortest distance from eye to otic vesicle in microns.

RESULTS

As compared to controls, copper treatment did not significantly change developmental timing as measured by morphological staging endpoints (Table 1). Individual milestones were reached at virtually the same time until hatch, where mean hatch occurred 6.18 hours later than the control mean (Figure 10). Likewise, TDS treatment did not significantly change developmental timing as measured by morphological staging endpoints. Morphological stages were reached at the same time as controls until hatch, where mean time to hatch

occurred 4 hours later than controls. No morphological abnormalities were observed in controls or either treatment.

Table 1. Time in minutes to reach each of the 7 previously identified morphological stages. Time start:hatch normalizes for start time of 64 cell stage.

Run number	Chemical	64-cell	Sphere	Shield	Bud	Eye/7 somites
nd027	Control	85	162	300	496	656
nd028	Control	85	148	321	543	691
nd029	Control	40	114	273	464	658
	Control mean	**70**	**141**	**298**	**501**	**668**
nd023	Copper	78	203	304	510	691
nd025	Copper	71	154	311	492	668
nd026	Copper	69	134	295	520	658
	Copper mean	**72**	**164**	**303**	**507**	**672**
nd019	TDS	76	145	272	478	627
nd022	TDS	85	161	316	514	671
nd016	TDS	77	143	330	587	758
	TDS mean	**79**	**149**	**306**	**526**	**685**

Table 1. Continued

Run number	Chemical	Movement	Otolith	Hatch	Time start:hatch	Total time in hours
nd027	Control	873	1095	2755	2670	
nd028	Control	925	1130	2904	2819	
nd029	Control	886	1132	2724	2684	
	Control mean	**894**	**1119**	**2794**	**2724**	45:40:00
nd023	Copper	913	1101	3343	3266	
nd025	Copper	893	1082	2946	2875	
nd026	Copper	872	1098	3212	3144	
	Copper mean	**893**	**1094**	**3167**	**3095**	51:58:00
nd019	TDS	883	1056	2927	2851	
nd022	TDS	909	1089	2792	2707	
nd016	TDS	1017	1183	3415	3338	
	TDS mean	**936**	**1109**	**3044**	**2965**	49:42:00

Comparison of slopes using morphological staging data revealed no significant delay in development for embryos treated with copper or TDS (Figure 10). The slope of control morphological staging data was set to 1 and the y-intercept to 0.

Treatment data used the same morphological staging y-axis as controls. Significant linear relationships existed for copper and TDS treatments ($p < 0.0001$, $R^2 = 0.99$). No statistically significant differences existed between control and treatment slopes ($F_{(1,2)} = 1.385$, $p = 0.2573$).

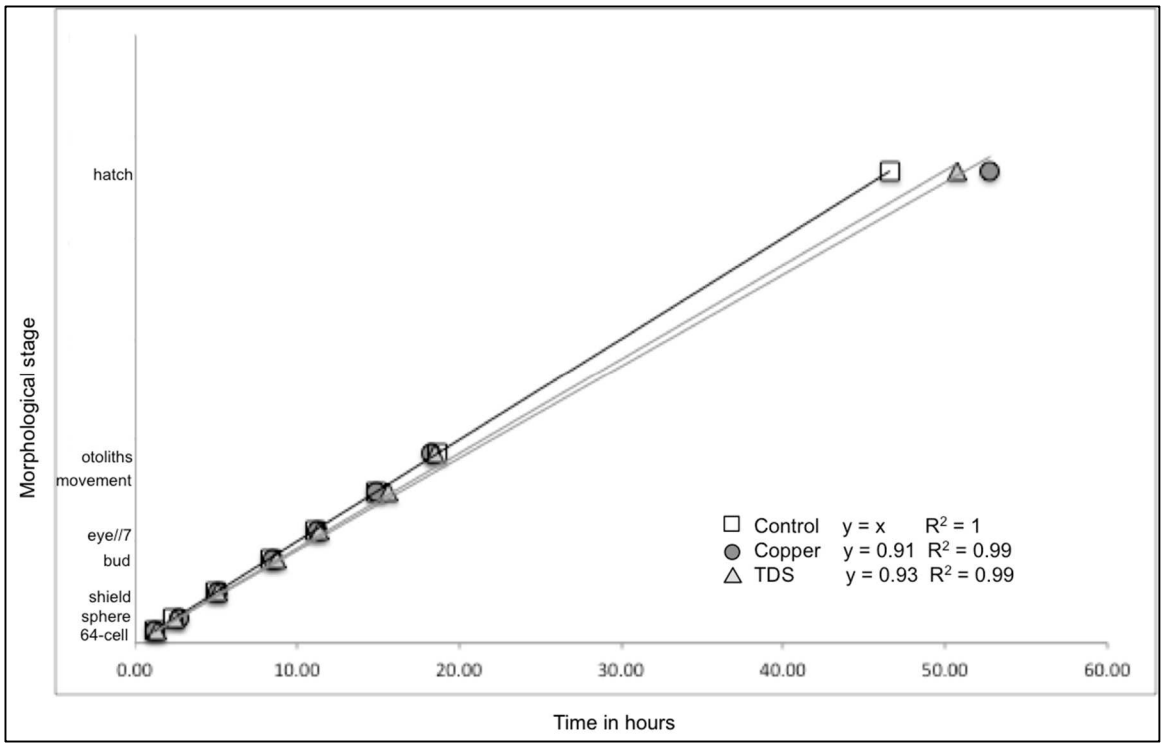

Figure 10. Mean time to reach morphological stages. $n = 3$ at each stage. Controls were set to a slope of 1 (y=x) as a reference line.

EED measurements revealed no significant delay in development for embryos treated with copper or TDS (Figure 11). A linear relationship between EED measurement and time existed for each treatment ($p < 0.0001$). Goodness of fit comparisons showed that EED slopes were precise (control $R^2 = 0.90$, copper

R^2 = 0.79, TDS R^2 = 0.68). Comparison of slopes of EED measurements using ANCOVA analysis showed no significant difference between rate of decrease in EED length in either copper or TDS as compared to controls ($F_{(1,2)}$ = 2.693, p = 0.0731). Y-intercepts were statistically different for all slopes (p < 0.0001).

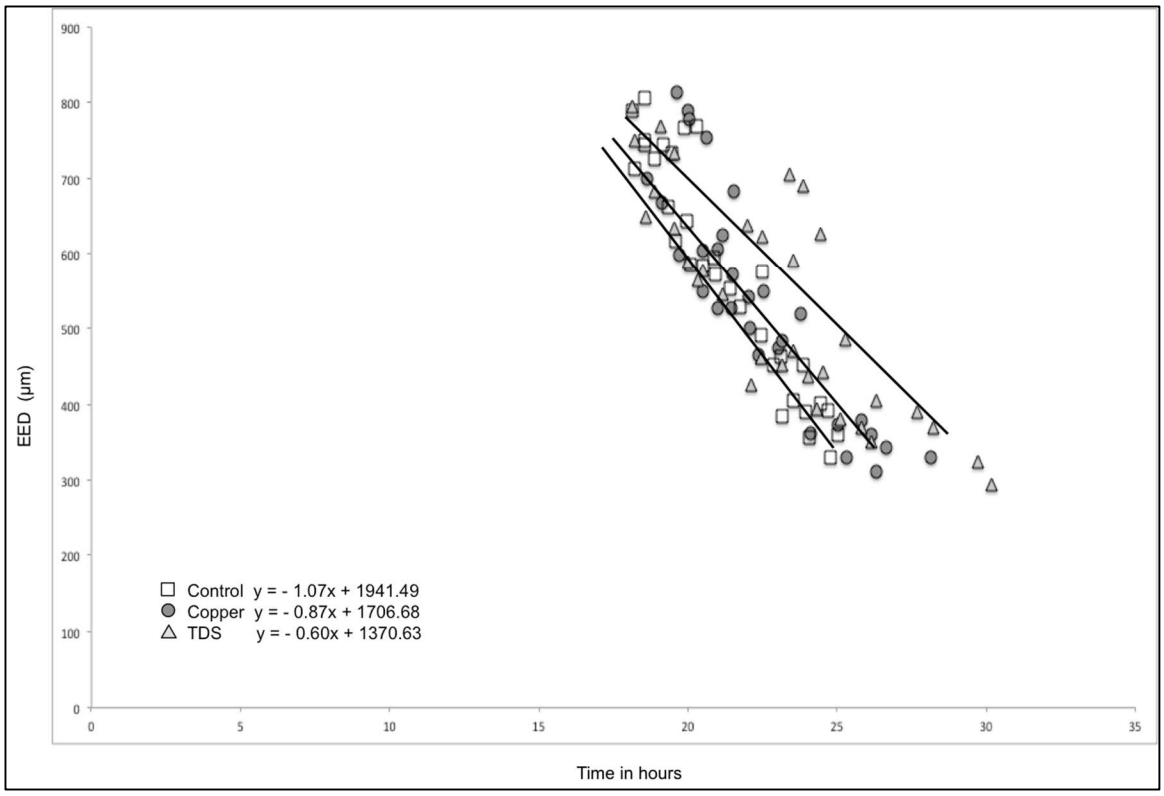

Figure 11. Ear-eye distance measurement. n = 21-25 for control and treatments.

DISCUSSION

With the OECD's introduction of the Fish Embryo Test (FET), the use of zebrafish embryos is increasingly common for assessment of acute effects. Such acute tests are designed to quickly measure mortality with the use of

several easily determined morphological endpoints (e.g., somite formation, hatch success) over a short timeframe (up to 96hpf). However, during this timeframe subtle changes at sublethal concentrations are also noteworthy as they may be indicative of effects that will only later become apparent. A common occurrence is development that appears to occur correctly without morphological abnormality but at a delayed rate. EDRA was developed as a refinement and simplification of previous work showing the effective use of combined morphological staging and EED measurements to track developmental rate [11].

Changes to developmental rate have been shown to be a useful biomarker of toxicity in embryogenesis, with morphological staging the most common technique to assess rate. Kimmel et al. [3] measured developmental rate over embryogenesis through the use of 18 morphological stages, Schirone and Gross [31] similarly with 16 stages, while Beasley et al. [11] used 13. In this study, a streamlined set of 7 morphological stages was used to cover the first 24hpf, when the majority of observable development occurs. After the 24-hour developmental window, movement of the fish precludes further identification of morphological stages without dechorionation and fixing. However, at approximately 20-24 hours, the otic vesicle and otolith (precursor to ear) become identifiable. Because of the angle of the developing embryo around the attached yolk sac, and the flattening of that angle as the fish develops and pulls away from the yolk sac, the distance between the eye and the otic vesicle shortens

throughout later development following a linear regression with a negative slope [3]. The rate of EED change is inversely associated with the linear rate of progression through developmental stages. EED was included in Kimmel's work as a secondary measure of developmental rate, while Beasley et al. compared EED with similar measurements to find that EED was a useful measurement method to confirm and further quantify morphological timing [3, 11]. In this study, EED was evaluated using 10-15 measurements starting at the initial appearance of the otolith. Special attention was paid to the plane of the fish, as it has been previously shown that the plane can profoundly affect the angle and therefore the measurement from ear to eye [11].

Embryos treated with copper and TDS, both well-characterized toxicants known to affect zebrafish development and hatching, were used to further demonstrate the utility of EDRA. Morphological staging showed no statistically significant difference between developmental rates in embryos treated with either test chemical as compared to controls. In treated embryos, the time to achieve six of the seven stages was nearly identical to that of controls. In addition, a goodness of fit comparison showed little variation between the three replicates of each treatment, both for copper and TDS ($p < 0.0001$, $R^2 = 0.99$ and $R^2 = 0.98$, respectively). The exception was the hatching endpoint, which due to natural biological variation was expected to vary [3, 32]. Between the three identical copper exposures, hatch time varied 4.48 hours, while TDS exposures' hatch

times occurred over a 10.51-hour span. Overall, the difference in mean time to hatch of either copper or TDS-exposed embryos was not statistically significant (Table 1, Figure 10). The variability in hatch time within exposure groups supports caution in using time to hatch as a standalone endpoint in measuring developmental delay. EED measurements for both treatments behaved similarly, with no significant difference in slopes between replicates and no significant effect due to treatment as compared to controls ($F_{(1,2)} = 2.693$, $p < 0.0731$). Interestingly, the minor delay in developmental rate shown in TDS treatments by hatch stage was echoed in the EED slopes, where the slope of copper treatment was closest to control slope and TDS treatment slightly delayed (Figure 11).

Although delays in developmental rate were not statistically significant in embryos exposed to TDS or copper sulfate, this was not inconsistent with expected effects of either compound. Both copper sulfate and TDS are known to primarily affect hatching success and timing; in this study, hatching success was affected in both positive controls and test embryos.

In this study, the Embryonic Developmental Rate Assay (EDRA) was created for use with zebrafish embryos to assess developmental rate as a biomarker of toxic effect. Morphological staging and EED measurement were combined to deliver a simple, effective method. EDRA was evaluated using copper and TDS, both known to affect rate of development in zebrafish embryos. Results of

morphological staging and EED measurement showed no significant difference in developmental rate at the chosen concentrations of each treatment. Morphological staging and EED delivered results consistent with each other in estimation of rate delay.

REFERENCES

[1] Scholz S, Sela E, Blaha L, Braunbeck T, Galay-Burgos M, Garcia-Franco M, Guinea J, Kluever N, Schirmer K, Tanneberger K. 2013. A European perspective on alternatives to animal testing for environmental hazard identification and risk assessment. *Regulatory toxicology and pharmacology* 67:506-530.

[2] OECD. 2013. *Test No. 236: Fish Embryo Acute Toxicity (FET) Test*. OECD Publishing.

[3] Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. 1995. Stages of embryonic development of the zebrafish. *Developmental dynamics* 203:253-310.

[4] Belanger SE, Balon EK, Rawlings JM. 2010. Saltatory ontogeny of fishes and sensitive early life stages for ecotoxicology tests. *Aquatic toxicology* 97:88-95.

[5] Strähle U, Scholz S, Geisler R, Greiner P, Hollert H, Rastegar S, Schumacher A, Selderslaghs I, Weiss C, Witters H. 2012. Zebrafish embryos as an alternative to animal experiments—a commentary on the definition of the onset of protected life stages in animal welfare regulations. *Reproductive toxicology* 33:128-132.

[6] Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch G-J, White S, Chow W, Kilian B, Quintais LT, Guerra-Assuncao JA, Zhou Y, Gu Y, Yen J, Vogel J-H, Eyre T, Redmond S, Banerjee R, Chi J, Fu B, Langley E, Maguire SF, Laird GK, Lloyd D, Kenyon E, Donaldson S, Sehra H, Almeida-King J, Loveland J, Trevanion S, Jones M, Quail M, Willey D, Hunt A, Burton J, Sims S, McLay K, Plumb B, Davis J, Clee C, Oliver K, Clark R, Riddle C, Eliott D, Threadgold G, Harden G, Ware D, Mortimer B, Kerry G, Heath P, Phillimore B, Tracey A, Corby N, Dunn M, Johnson C, Wood J, Clark S, Pelan S, Griffiths G, Smith M, Glithero R, Howden P, Barker N, Stevens C, Harley J, Holt K, Panagiotidis G, Lovell J, Beasley H, Henderson C, Gordon D, Auger K, Wright D, Collins J, Raisen C, Dyer L, Leung K, Robertson L, Ambridge K, Leongamornlert D, McGuire S, Gilderthorp R, Griffiths C, Manthravadi D, Nichol S, Barker G, Whitehead S, Kay M, Brown J, Murnane C, Gray E, Humphries M, Sycamore N, Barker D, Saunders D, Wallis J, Babbage A, Hammond S, Mashreghi-Mohammadi M, Barr L, Martin S, Wray P, Ellington A, Matthews N, Ellwood M, Woodmansey R, Clark G, Cooper J, Tromans A, Grafham D, Skuce C, Pandian R, Andrews R, Harrison E, Kimberley A, Garnett J, Fosker N, Hall R, Garner P, Kelly D, Bird C, Palmer S, Gehring I, Berger A, Dooley CM, Ersan-Urun Z, Eser C, Geiger H, Geisler M, Karotki L, Kirn A, Konantz J, Konantz M, Oberlander M, Rudolph-Geiger S, Teucke M, Osoegawa K, Zhu B, Rapp A, Widaa S, Langford C, Yang F, Carter NP, Harrow J, Ning Z, Herrero J, Searle SMJ, Enright A, Geisler R, Plasterk RHA, Lee C,

Westerfield M, de Jong PJ, Zon LI, Postlethwait JH, Nusslein-Volhard C, Hubbard TJP, Crollius HR, Rogers J, Stemple DL. 2013. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496:498-503.
[7] Brion F, Le Page Y, Piccini B, Cardoso O, Tong S-K, Chung B-c, Kah O. 2012. Screening estrogenic activities of chemicals or mixtures in vivo using transgenic (cyp19a1b-GFP) zebrafish embryos. *PloS one* 7:e36069.
[8] OECD. 1992. *Test No. 203: Fish, Acute Toxicity Test*. OECD Publishing.
[9] Bowman CR, Bailey FC, Elrod-Erickson M, Neigh AM, Otter RR. 2012. Effects of silver nanoparticles on zebrafish (Danio rerio) and Escherichia coli (ATCC 25922): a comparison of toxicity based on total surface area versus mass concentration of particles in a model eukaryotic and prokaryotic system. *Environmental toxicology and chemistry* 31:1793-1800.
[10] Lammer E, Carr G, Wendler K, Rawlings J, Belanger S, Braunbeck T. 2009. Is the fish embryo toxicity test (FET) with the zebrafish (< i> Danio rerio</i>) a potential alternative for the fish acute toxicity test? *Comparative biochemistry and physiology part C: toxicology & pharmacology* 149:196-209.
[11] Beasley A, Elrod-Erickson M, Otter RR. 2012. Consistency of morphological endpoints used to assess developmental timing in zebrafish (*Danio rerio*) across a temperature gradient. *Reproductive toxicology* 34:561-567.
[12] Hallare A, Schirling M, Luckenbach T, Köhler H-R, Triebskorn R. 2005. Combined effects of temperature and cadmium on developmental parameters and biomarker responses in zebrafish (*Danio rerio*) embryos. *Journal of thermal biology* 30:7-17.
[13] Bar-Ilan O, Albrecht RM, Fako VE, Furgeson DY. 2009. Toxicity assessments of multisized gold and silver nanoparticles in zebrafish embryos. *Small* 5:1897-1910.
[14] Walpita CN, Crawford AD, Janssens ED, Van der Geyten S, Darras VM. 2009. Type 2 iodothyronine deiodinase is essential for thyroid hormone-dependent embryonic development and pigmentation in zebrafish. *Endocrinology* 150:530-539.
[15] Beasley A, Graham C, Otter R, Elrod-Erickson M. 2014. A molecular method for assessing the effects of potential contaminants on the rate of zebrafish (*Danio rerio*) development. *Environmental toxicology and chemistry* 33:238-242.
[16] Peravali R, Gehrig J, Giselbrecht S, Lütjohann DS, Hadzhiev Y, Müller F, Liebel U. 2011. Automated feature detection and imaging for high-resolution screening of zebrafish embryos. *Biotechniques* 50:319.
[17] Letamendia A, Quevedo C, Ibarbia I, Virto JM, Holgado O, Diez M, Belmonte JCI, Callol-Massot C. 2012. Development and validation of an automated high-throughput system for zebrafish in vivo screenings. *PloS one* 7:e36690.

[18] Belanger SE, Rawlings JM, Carr GJ. 2013. Use of fish embryo toxicity tests for the prediction of acute fish toxicity to chemicals. *Environmental toxicology and chemistry* 32:1768-1783.
[19] McKim J, Eaton J, Holcombe GW. 1978. Metal toxicity to embryos and larvae of eight species of freshwater fish—II: copper. *Bulletin of environmental contamination and toxicology* 19:608-616.
[20] Woltering DM. 1984. The growth response in fish chronic and early life stage toxicity tests: a critical review. *Aquatic toxicology* 5:1-21.
[21] Johnson A, Carew E, Sloman K. 2007. The effects of copper on the morphological and functional development of zebrafish embryos. *Aquatic toxicology* 84:431-438.
[22] Dowse R, Tang D, Palmer CG, Kefford BJ. 2013. Risk assessment using the species sensitivity distribution method: data quality versus data quantity. *Environmental toxicology and chemistry* 32:1360-1369.
[23] Westernhagen Hv, Dethlefsen V, Rosenthal H. 1979. Combined effects of cadmium, copper and lead on developing herring eggs and larvae. *Helgoländer wissenschaftliche meeresuntersuchungen* 32:257-278.
[24] Iwase H, Suga S, Shimada M, Yamada H, Horiuchi Y, Oohashi M. 1996. Eleven-year survey of safety and efficacy of endoscopic injection sclerotherapy using 2% sodium tetradecyl sulfate and contrast medium. *Journal of clinical gastroenterology* 22:58-65.
[25] Versteeg DJ, Stanton DT, Pence MA, Cowan C. 1997. Effects of surfactants on the rotifer, Brachionus calyciflorus, in a chronic toxicity test and in the development of QSARs. *Environmental toxicology and chemistry* 16:1051-1058.
[26] Belanger SE, Lee DM, Bowling JW, LeBlanc EM. 2004. Responses of periphyton and invertebrates to a tetradecyl-pentadecyl sulfate mixture in stream mesocosms. *Environmental toxicology and chemistry* 23:2202-2213.
[27] Busquet F, Strecker R, Rawlings JM, Belanger SE, Braunbeck T, Carr GJ, Cenijn P, Fochtman P, Gourmelon A, Hübler N. 2014. OECD validation study to assess intra-and inter-laboratory reproducibility of the zebrafish embryo toxicity test for acute aquatic toxicity testing. *Regulatory toxicology and pharmacology*. 69(3):496-511
[28] Belanger S, Rupe K, Bausch R. 1995. Responses of invertebrates and fish to alkyl sulfate and alkyl ethoxylate sulfate anionic surfactants during chronic exposure. *Bulletin of environmental contamination and toxicology* 55:751-758.
[29] Lizotte Jr R, Wong D, Dorn P, Rodgers Jr J. 1999. Effects of a homologous series of linear alcohol ethoxylate surfactants on fathead minnow early life stages. *Archives of environmental contamination and toxicology* 37:536-541.
[30] OECD. 2012. *Validation Report (Phase 2) for the Zebrafish Embryo Toxicity Test (Annexes) No. 179*. OECD Publishing.

[31] Schirone RC, Gross L. 1968. Effect of temperature on early embryological development of the zebra fish, Brachydanio rerio. *Journal of experimental zoology* 169:43-52.
[32] Fraysse B, Mons R, Garric J. 2006. Development of a zebrafish 4-day embryo-larval bioassay to assess toxicity of chemicals. *Ecotoxicology and environmental safety* 63:253-267.

CHAPTER II

Stepwise Information-Filtering Tool (SIFT): A method for using risk assessment metadata in a nontraditional way

INTRODUCTION

'Big data' just keeps getting bigger. The increasing accessibility to data and ease of collection, plus new regulatory requirements under international chemical management programs such as REACH, OECD High Production Volume Challenge, and Categorization of the Canadian Domestic Substances List have resulted in massive ecotoxicology metadatasets [1]. Chemical registration through the European Chemical Agency (ECHA) has produced a database of dossier information on the toxicity of over 12,000 chemical substances manufactured at over ten tons a year [2]. Others, like the United States Environmental Protection Agency (USEPA)'s ECOTOX database of more than 10,000 individual chemicals, are a collection of toxicity test information from government, peer-reviewed literature and private sources [3]. These datasets can be thought of as massive data warehouses in which the data submitted or gathered is all directed towards eventual use in risk assessment methodologies. Similar datasets on a smaller scale are compiled by private industry and government towards the same risk assessment endpoints, but may be tailored to a subset of chemicals (e.g. pesticides, pharmaceuticals) [4, 5] or housed by a particular organization (e.g. CAL-ECOTOX, Columbia ERC) [6, 7].

In typical risk assessment methodology, database size can be an advantage. A traditional chemical risk assessment workflow begins with gathering all available toxicity data, since more toxicity data equals more certainty about potential risk. Ideally, the dataset would incorporate multiple trophic levels and multiple toxicity endpoints. Upon data evaluation for suitability an assessment of acute and chronic toxicity endpoints would be used to derive a Predicted No Effect Concentration (PNEC). Uncertainty factors are applied to sensitive endpoints based on the depth and breadth of available data. Uncertainty factors vary by region or regulatory authority based on considerations of data requirements in each jurisdiction [2, 8].

Reliability of the data used to derive the final PNEC is an important consideration; reliability can be determined by evaluating the confidence in test methodology used, adherence to that methodology, and accurate reporting of the test event. REACH and ECOTOX databases rely on specific submission guidelines to ensure that data quality is represented accurately [2]. Tools exist that are useful for data evaluation in a traditional hazard/risk assessment context [9-11] and that use a set of predefined criteria to evaluate data reliability. Klimisch (1997) developed a simple, well-known scoring method for assessing reliability based on adherence to sound science, quality assurance, and Good Laboratory Practice (GLP) guidelines [19]. The ToxRTool is a Klimisch-based tool intended to expand and clarify definitions of reliability for use in REACH registration compliance [12]. Similar tools are designed to evaluate specific chemical types

(e.g., nanomaterials, pesticides) [4, 13] and data types (e.g., nonstandard toxicity data)[14-16]. Agerstrand and colleagues created a set of comprehensive criteria that uses relevance as well as reliability criteria for use in evaluating pharmaceuticals [5]. Although these tools are different from one another, each serves a similar purpose, which is to evaluate data toward a single hazard/risk endpoint target.

Given the size and complexity of risk assessment datasets, it seems obvious that such a collection could be mined many ways to serve many purposes. However, size and complexity can become a disadvantage without a method to categorize and select the best data for the purpose [17]. Data gaps and overlaps are likely, and data quality may be difficult to quantify [11, 18]. This study was designed to 1) develop a strategic and systematic method using user-defined criteria to evaluate large datasets then 2) demonstrate the usefulness of this method in a case study of chronic toxicity test reports for the purpose of evaluating improvements to test methodology and 3) further evaluate the unique function of the methodology by comparing case study results to both the Klimisch and Agerstrand methods [5, 19].

METHODS

Development of SIFT

The Stepwise Information-Filtering Tool (SIFT) was developed to evaluate and refine large datasets, with an emphasis on data relevance and reliability. Initial

user-defined data selection criteria used in the development of SIFT were based on, EPA test guidelines, OECD test guidelines, and expert judgment (Figure 1) [20-24]. It is important to note that although the order of the steps in the SIFT methodology remains the same for each user/purpose, the criteria within each step are completely user-defined; for instance, the Step 2 validity criteria would be dependent upon the relevant test guidelines selected in Step 1.

SIFT Step 0 - **Define the Dataset**. A purpose for the study is defined. A broad master dataset is compiled that generally covers the defined purpose.

Step 1: **Relevance Criteria**. Relevance criteria narrow the master dataset based on the defined purpose.

Step 2: **Validity Criteria**. Validity criteria, including, but not limited to, toxicity test guidelines narrow the Step 1 dataset.

Step 3: **Acceptability Criteria**. Acceptability criteria described by desired parameters of test design and reporting narrow the Step 2 dataset.

Step 4: **Additional Criteria**. Additional user-defined criteria relevant to the defined purpose narrow the Step 3 dataset to the final set of studies.

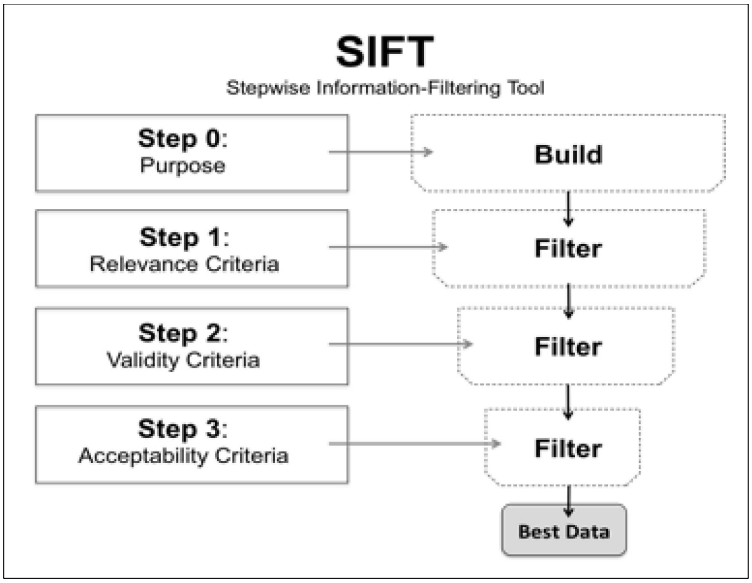

Figure 1. Process diagram of the SIFT methodology.

Case Study

In order to demonstrate the functionality of SIFT a case study was conducted based on a company's request for comparative analysis of the statistical measurement endpoints in chronic invertebrate toxicity test reports. Criteria specific to this purpose were identified for each of the SIFT steps (Table 1). The master dataset was compiled and then narrowed to a final dataset using the identified SIFT criteria.

Table 1. SIFT criteria developed specific to the purpose of the endpoint comparison case study.

Step 1: Relevance criteria
Tests conducted under OECD Daphnia reproduction test protocol TG211
Species (*Daphnia magna* or *Ceriodaphnia dubia*)
Reported NOEC value
Effect data available to the level of individual replicate
Step 2: Validity criteria
≤ 20% control mortality
Minimum of 60 young per surviving adult by end of test
Condition of organism (first instar, no ephippia, from the same culture)
Test duration
Step 3: Acceptability criteria
Actual concentrations reported
Minimum 5 concentrations plus control
Parameters included or available: test type, test strategy, chemical class, solvent used
CAS number when available
Dose response present
No hormetic effect
Step 4: Additional criteria
Single compound
Two of the following criteria reported:
Hardness
pH
Kow
Temperature

Methodology Comparison

User-defined SIFT criteria as utilized in the endpoint comparison case study and criteria from methods described by Klimisch et al. [19] and Agerstrand et al. [5] were used to evaluate the dataset for the defined case study purpose (Table 2,3). Agerstrand lists criteria separated into Relevance and Reliability then notes each as mandatory or optional. SIFT criteria used in the case study were

compared to the Agerstrand criteria for relevance and reliability then noted as mandatory, not useful (or necessary), or unclear.

Table 2. Comparison of Agerstrand versus SIFT criteria as applied to the case study dataset. O denotes criteria standard to OECD TG211, x denotes criteria utilized with definitions as stated by Agerstrand, and * denotes criteria definitions with caveats (e.g., Agerstrand's strain and clone are mandatory for algal tests and *Daphnia* only). Relevance criteria.

Agerstrand Criteria	**Application to Case Study**				
Relevance Criteria	**Agerstrand Mandatory**	**Agerstrand Optional**	**SIFT Mandatory**	**SIFT Not Useful**	**SIFT Criteria Unclear**
Is the substance tested representative for the substance being risk assessed?		x		x	
Is the appropriate test species studied		x	O		
Are the appropriate life-stage(s) studied?		x	O		
Are the appropriate endpoint(s) studied?		x	x		
Is the route of exposure relevant for the species?		x	O		
Does the test exposure scenario exist for the tested substance?		x			x
Are the stated tested doses/concentrations appropriate?		x			x
How do the tested doses relate to measured or predicted environmental concentrations (if available)?		x		x	
Is the time of exposure relevant and appropriate for the studied endpoints?		x	O		
Have other critical parameters influencing the endpoints than exposure time been considered adequately?		x	x		
Should the measured endpoint be considered to be an adverse effect or not?		x		x	
Are the references reported?	x		x		

Table 2 continued. Reliability criteria.

Agerstrand Criteria	Application to Case Study				
Reliability Criteria	Agerstrand Mandatory	Agerstrand Optional	SIFT Mandatory	SIFT Not Useful	SIFT Criteria Unclear
Purpose of study		x		x	
Description of endpoints	x		O		
Protocol Standard/modified standard (if used)	x		O		
Test compound Identification (e.g. name, CAS-number)	x		x*		
Physico-chemical data (e.g. volatility, stability,solubility, degradability, adsorption)		x		x	
Source		x		x	
Purity	x			x	
Vehicle (if used)	x		O		
Radiolabelled (if used)		x		x	
Tested doses or concentrations	x		O		
Measured doses or concentrations.	x		x		
Exposure duration	x		O		
Exposure route	x		x*		
Exposure schedule (static, semistatic, flow through system, other)	x		x		
Method of preparation of stock solutions		x		x	
Time-points of observations	x		O		
Analytical method	x			x	
Scientific name	x		O		
Body weight or length		x	x*		
Age/life-stage	x		O		
Growth/reproductive condition		x	O		
Gender	x*	x*		x	
Strain, clone	x*	x*	O		
Source	x			x	
Culture handling	x			x	
History of contamination for field-collected species	x			x	
Control described	x		O		
Control media identical to test media in all respect except treatment	x		O		
Control(s) identical to treatments in physical and chemical test conditions aspects: light, location, temperatu	x		O		
Control and test organism drawn from same population	x		O		
Control mortality/morbidity	x		O		
Positive/negative control (if used)	x		O		
Vehicle control (if used)	x		O		
Known concentrations of vehicle (if used) in treatments and control	x		O		
Control mortality/morbidity reported for vehicle/positive control (if used))	x		O		
Historical control data		x		x	
Test environment pH	x		x		
Temperature	x		x		
Conductivity	x*	x*		x	
Light intensity and quality (source and homogeneity)	x*	x*		x	
Photo period	x			x	
Hardness of water		x	x		
Dissolved oxygen content	x*			x	
Ammonium/nitrite content in water	x*	x*		x	
Material and volume on aquarium/container	x		O		
Test medium	x		O		
Feeding protocols (for long-term tests)	x		O		
Food composition		x		x	
Sample size per replicates, number of organisms per replicates	x		O		
No of organisms from each replicates used for statistical analysis (if not all used)	x				x
Randomized treatments		x	O		
Independence of observations		x	O		
Statistical method used	x				x
Significance level for NOEC and LOEC data (0.05 or less)	x		x		
Estimate of variability for LC and EC data	x			x	
Results reproduced by others		x		x	
Consistent with other findings		x		x	
Statistically significant responses noted (e.g. ECx)	x			x	
Dose-response reported in figure/text/table	x			x	
Each effect concentrations explicitly related to a specific endpoint		x			x
References to support the reliability of the study should be reported		x			x
Produced according to GLP		x		x	
Availability of raw data		x	x		

Table 3. Criteria developed by Klimisch et al. (1997) to evaluate reliability. Figure adapted from Schneider et al. (2009)[12,18].

Category	Definition
1. Reliable without restrictions	"Studies or data from the literature or reports which were carried out or generated according to generally valid and/or internationally accepted testing guidelines (preferably performed according to GLP) or in which the test parameters documented are based on a specific (national) testing guideline (preferably performed according to GLP) or in which all parameters described are closely related/comparable to a guideline method."
2. Reliable with restrictions	"Studies or data from the literature, reports (mostly not performed according to GLP), in which the test parameters documented do not totally comply with the specific testing guideline, but are sufficient to accept the data or in which investigations are described which cannot be subsumed under a testing guideline, but which are nevertheless well documented and scientifically acceptable."
3. Not reliable	"Studies or data from the literature/reports in which there were interferences between the measuring system and the test substance or in which organisms/test systems were used which are not relevant in relation to the exposure (e.g., unphysiologic pathways of application) or which were carried out or generated according to a method which is not acceptable, the documentation of which is not sufficient for assessment and which is not convincing for an expert judgment."
4. Not assignable	"Studies or data from the literature, which do not give sufficient experimental details and which are only listed in short abstracts or secondary literature (books, reviews, etc.)."

RESULTS

Case study

Step 0 – Define the Dataset

The purpose of the case study was defined as ‘a comparative analysis of the statistical measurement endpoints in chronic invertebrate toxicity test reports’. Database construction was initiated with a company-provided archive of reports spanning over 30 years and included toxicity information on a wide range of chemicals. A targeted search of peer-reviewed literature was also performed to supplement the provided archive. The master dataset included **210** studies.

Step 1 – Relevance Criteria

Based on user-defined Relevance Criteria (Table 1) that fit the purpose identified in Step 0, the master dataset was narrowed from **210** to **156** studies. Details of OECD TG211 can be found in supplemental table 1. Although studies from the literature rarely included effect data to the individual replicate level, two studies did pass the Step 1 criteria. The NOEC was defined as the No Observed Effect Concentration. The NOEC was reported or, when only a LOEC (Lowest Observed Effect Concentration) was reported, the NOEC was determined to be the next lowest test concentration from the LOEC.

Step 2 - Validity Criteria

Based on user-defined Validity Criteria (Table 1), the Step 1 dataset was narrowed from **156** to **136** studies. Control mortality was defined per OECD TG211 for both species and minimum young applied to *Daphnia magna*. Valid study duration, in the context of an OECD TG211 for *Daphnia magna* and USEPA Method 1002.0 for *Ceriodaphnia dubia* was defined as 21-23 days and 7-8 days, respectively.

Step 3 - Acceptability Criteria

Based on user-defined Acceptability Criteria (Table 1), the dataset from Step 2 was narrowed from **136** to **95** studies. Effect concentrations for each study were calculated from reported raw effect data. Test type included semi-static and flow-through. Test strategy referred to configuration of replicates per concentration (e.g., 10 replicates of 1 daphnid, 4 replicates of 5 daphnid). At this step, chemical class was used for categorization of the database since some chemicals in the archive were either not in production or were proprietary formulations without CAS numbers. CAS numbers were used to confirm class categorization when available. Exposure response and hormesis were analyzed from reported raw effect data.

Step 4 - Additional Criteria

Based on user-defined Additional Criteria (Table 1), the Step 3 dataset was narrowed from **95** to **78** studies. Only tests of single-compound chemicals were

included. Two of the following four parameters had to be reported for inclusion in the dataset: temperature, hardness, Kow, pH.

Methodology Comparison

SIFT v. Agerstrand

Agerstrand's method included 75 total criteria. Of 12 relevance criteria, 11 were optional and only one mandatory (reporting of references). Of 63 reliability criteria, 46 were mandatory and 23 optional, with many criteria represented in both mandatory and optional depending on the response given (Table 2). Using criteria specifically defined for the case study, SIFT would have considered 4 of Agerstrand's relevance criteria as mandatory per OECD TG211. Three additional relevance criteria would have been mandatory, while 3 were not useful and 2 unclear. Of the 63 reliability criteria in Agerstrand's method, 25 would have been mandatory with SIFT per OECD TG211. Ten criteria would have been otherwise mandatory, while 24 would have been defined not useful (or not necessary) and 4 unclear. SIFT did not treat any criteria as optional.

SIFT v. Klimisch

Klimisch's method included 4 reliability criteria. In the case study, all user-defined criteria were largely based on the OECD TG211 as specified under Step 1, which would have resulted in a Klimisch score of 1 or 2 depending on expert judgment of whether test guidelines were followed appropriately. Klimisch does not include relevance criteria (Table 3).

DISCUSSION

Complex ecotoxicology datasets are becoming more common and more accessible. With such 'big data', evaluation of data reliability is important. Although there are tools to evaluate data in a risk assessment context, gaps still exist for tools to evaluate the same data for other approaches. The Klimisch et al. (1997) method was intended to simplify and clarify the data evaluation process and were considered highly useful for uses associated with impending REACH legislation[19]. With the Klimisch method, data is generally categorized into one of four reliability classes in order to include the maximum reliable data to build a weight of evidence approach. The flexibility of the category definitions allows the method to be tailored to numerous risk assessment scenarios (e.g., *in vitro* and *in vivo* testing). Critics claim that the broad categories are too open to interpretation, thus complicating standardization and transparency; another concern is the expertise needed to correctly interpret and apply evaluation criteria [25-27]. Nevertheless, the Klimisch method is the most frequently cited method to date (Web of Science™ Core Collection) and is now part of the ECHA guidance on REACH registration [28]. Agerstrand et al. (2011) created a more comprehensive method for evaluation of pharmaceutical data in risk assessment [5]. To counter the typical focus on reliability, this method defines relevance criteria and emphasizes clear, comprehensive criteria definitions for increased transparency and ease of use; however, such explicit definitions can result in a narrow range of application [16]. The Klimisch and Agerstrand methods were

chosen to compare to SIFT because they were representative of a variety of existing tools.

Relevance

When the purpose of a study is a comparison within a dataset as opposed to the traditional comparison to an external reference value, ‘relevance’ becomes about initial data selection. This shift in approach means that data relevance is completely divorced from data reliability. SIFT uses upfront data-collection decisions that apply directly to each particular purpose. Data that do not include parameters key to the purpose are not useful, regardless of reliability. In an example from the case study, the presence of raw effect data to the individual replicate was essential to final endpoint calculations and comparisons. Therefore ‘effect data present’ was chosen as a SIFT Step 1 criterion in the case study; Klimisch did not specify criteria that would address raw data, whereas Agerstrand did but as an optional, late-stage criterion.

Both Klimisch and Agerstrand consider relevance to be dependent on reliability. The Klimisch method does not predefine criteria to establish relevance. Instead, the method provides general guidance on how to use relevance to compare equally reliable tests. The Agerstrand method elaborates on this approach, citing the influence of REACH guidance to assess relevance as “appropriateness of the test when it comes to a particular risk, e.g. whether the experimental model is representative to the environment that is aimed to be protected”[5]. Unlike

Klimisch, Agerstrand's method includes predefined relevance criteria, and relevance is considered equally as important as reliability to a complete evaluation of data quality. In the case study, the selected Step 1 relevance criteria netted 210 studies from proprietary and public literature that were suitable for further reliability screening. Klimisch's method would have required that any and all invertebrate studies available in the proprietary database be screened for validity since the original purpose did not specify a particular test guideline or species. Agerstrand's method would not have allowed 3 of the 4 case study Step 1 criteria as relevance criteria and would have introduced 11 additional criteria to determine relevance. SIFT's focus on relevance in the initial data selection means that subsequent evaluations for reliability and acceptability are targeted only to useful data.

Balance between flexibility and utility

The development of SIFT stemmed from the difficulty in interpretation and adaptation of existing methods to the case study purpose. Because the initial question was not a traditional hazard/risk assessment and the case study was conducted by a research scientist rather than a risk assessment expert, both ease of use and flexibility in the method were important. Similar difficulties even within the risk assessment community have led to refinement or creation of new methods [9, 11, 29]. It seems that criteria definitions are the crux of the balance between flexibility to tailor the method and the ability to use the method correctly. Here, Klimisch and Agerstrand represent opposing ends of the spectrum;

Klimisch uses a few broad categories that require expert interpretation, while Agerstrand uses many specific criteria, partly to alleviate the need for expert interpretation. If criteria are thoroughly defined, transparency of the data and the possibility of standardization are increased [5]. Specific definitions also minimize the need for expert judgment; if criteria are spelled out, then less experienced users can confidently assess data quality and the likelihood of consistent data evaluation rises [9]. Conversely, a method with many narrow criteria may appear so complex and cumbersome that it is not utilized [17, 30]. Maybe more importantly, data that might otherwise be useful could be discarded, therefore diminishing the sample size of data necessary to decision-making.

SIFT is intended to preserve flexibility through broadly defined steps and upfront selection of criteria specific to the end-goal analysis. SIFT will still benefit from expert judgment; however, a basic understanding of toxicity test framework should allow the systematic breakdown of a dataset using SIFT criteria.

Application

SIFT is fundamentally about versatility in application. Both Klimisch and Agerstrand methods were created to serve very specific purposes: Klimisch to broadly characterize data quality for inclusion in chemical registration databases, and Agerstrand to improve reliability and harmonization of data reporting in scientific literature with an emphasis on pharmaceuticals.

The goal of SIFT is to make risk assessment metadatasets more accessible to more applications: expert/non-expert, basic/applied science, industry/government/academia. As part of the toxicity dataset analysis process, SIFT would be useful to inform such business decisions as the evaluation of contract lab performance or cost basis of in-house testing. Other potential applications include studies similar to Dowse et al. (2013) or Euling et al. (2013). Dowse et al. (2013) used toxicity data to compare standard testing to rapid testing in the construction of species sensitivity distributions [15]. Euling et al. (2013) performed a case study of toxicity data for dibutyl phthalate exposure to determine whether toxicogenomic data would be useful to elucidating modes of action [31].

CONCLUSION

SIFT is a useful addition to existing methods for risk assessment data evaluation. Current methods aim to evaluate large toxicity datasets toward traditional risk assessment purposes, leaving a gap for a method to evaluate these datasets nontraditionally. SIFT applies user-defined evaluation criteria that gauge relevance and reliability in a stepwise manner, keeping a balance between flexibility and utility of the method. SIFT is applicable across disciplines and levels of expertise.

REFERENCES

[1] European Parliament, The Council of the European Union. 2006. Regulation (EC) no. 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the registration, evaluation, authorization and restriction of chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) no. 793/93 and Commission Regulation (EC) no. 1488/94, as well as Council Directive 76/769/EEC and commission Directives 91/155/EEC, 93/67/EEC, 93/105/CE and 2000/21/EC. *Official journal of the European Union* 396.

[2] ECHA (European Chemicals Agency). 2008. Chapter R.7b: Endpoint specific guidance. *Guidance on information requirements and chemical safety assessment* Vol May 2008, p. 65.

[3] U.S. Environmental Protection Agency. 2012. ECOTOX User Guide: ECOTOXicology Database System. Version 4.0. Washington, DC.

[4] Roncaglioni A, Benfenati E, Boriani E, Clook M. 2004. A protocol to select high quality datasets of ecotoxicity values for pesticides. *Journal of environmental science and health, part B* 39:641-652.

[5] Ågerstrand M, Küster A, Bachmann J, Breitholtz M, Ebert I, Rechenberg B, Ruden C. 2011. Reporting and evaluation criteria as means towards a transparent use of ecotoxicity data for environmental risk assessment of pharmaceuticals. *Environmental pollution* 159:2487-2492.

[6] Office of Environmental Health Hazard Assessment (OEHHA) State of California, University of California at Davis. 2003. California Wildlife Biology, Exposure Factor, and Toxicity Database (Cal/Ecotox) State of California.

[7] U.S. Department of the Interior, U.S. Geological Survey. 2001. Columbia Environmental Research Center Acute Toxicity Database. Columbia, MO.

[8] Zeeman M, Fairbrother A, Gorsuch JW. 1995. Environmental toxicology: testing and screening. *Screening and testing chemicals in commerce* 23:32.

[9] Hobbs DA, Warne MSJ, Markich SJ. 2005. Evaluation of criteria used to assess the quality of aquatic toxicity data. *Integrated environmental assessment and management* 1:174-180.

[10] Breton RL, Gilron G, Thompson R, Rodney S, Teed S. 2009. A new quality assurance system for the evaluation of ecotoxicity studies submitted under the new substances notification regulations in Canada. *Integrated environmental assessment and management* 5:127-137.

[11] Durda JL, Preziosi DV. 2000. Data quality evaluation of toxicological studies used to derive ecotoxicological benchmarks. *Human and ecological risk assessment* 6:747-765.

[12] Schneider K, Schwarz M, Burkholder I, Kopp-Schneider A, Edler L, Kinsner-Ovaskainen A, Hartung T, Hoffmann S. 2009. “ToxRTool”, a new tool to assess the reliability of toxicological data. *Toxicology letters* 189:138-144.

[13] Card JW, Magnuson BA. 2010. A method to assess the quality of studies that examine the toxicity of engineered nanomaterials. *International journal of toxicology* 29:402-410.

[14] Parkerton TF, Arnot JA, Weisbrod AV, Russom C, Hoke RA, Woodburn K, Traas T, Bonnell M, Burkhard LP, Lampi MA. 2008. Guidance for evaluating in vivo fish bioaccumulation data. *Integrated environmental assessment and management* 4:139-155.
[15] Dowse R, Tang D, Palmer CG, Kefford BJ. 2013. Risk assessment using the species sensitivity distribution method: data quality versus data quantity. *Environmental toxicology and chemistry* 32:1360-1369.
[16] Ågerstrand M, Breitholtz M, Rudén C. 2011. Comparison of four different methods for reliability evaluation of ecotoxicity data: a case study of non-standard test data used in environmental risk assessments of pharmaceutical substances. *Environmental sciences Europe* 23:1-15.
[17] Przybylak K, Madden J, Cronin M, Hewitt M. 2012. Assessing toxicological data quality: basic principles, existing schemes and current limitations. *SAR and QSAR in environmental research* 23:435-459.
[18] Wheeler J, Grist E, Leung K, Morritt D, Crane M. 2002. Species sensitivity distributions: data and model choice. *Marine pollution bulletin* 45:192-202.
[19] Klimisch H-J, Andreae M, Tillmann U. 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regulatory toxicology and pharmacology* 25:1-5.
[20] OECD. 2004. *Test No. 202: Daphnia sp. Acute Immobilisation Test*. OECD Publishing.
[21] OECD. 2012. *Test No. 211: Daphnia magna Reproduction Test*. OECD Publishing.
[22] OECD. 2013. *Test No. 236: Fish Embryo Acute Toxicity (FET) Test*. OECD Publishing.
[23] OECD. 2011. *Test No. 201: Freshwater Alga and Cyanobacteria, Growth Inhibition Test*. OECD Publishing.
[24] OECD. 1992. *Test No. 210: Fish, Early-Life Stage Toxicity Test*. OECD Publishing.
[25] Maxim L, Van der Sluijs JP. 2014. Qualichem In Vivo: A tool for assessing the quality of in vivo studies and its application for bisphenol A. *PloS one* 9:e87738.
[26] Petersen K, Lindeman B. 2014. *Road to regulation of endocrine disruptors and combination effects*. Nordic Council of Ministers.
[27] de Vries P, Murk AJ. 2013. Compliance of LC50 and NOEC data with Benford's Law: An indication of reliability? *Ecotoxicology and environmental safety* 98:171-178.
[28] ECHA (European Chemicals Agency). 2012. *Guidance on data sharing, version 20*. Vol April 2012, p. 148.
[29] Küster A, Bachmann J, Brandt U, Ebert I, Hickmann S, Klein-Goedicke J, Maack G, Schmitz S, Thumm E, Rechenberg B. 2009. Regulatory demands on data quality for the environmental risk assessment of pharmaceuticals. *Regulatory toxicology and pharmacology* 55:276-280.

[30] Money CD, Tomenson JA, Penman MG, Boogaard PJ, Jeffrey Lewis R. 2013. A systematic approach for evaluating and scoring human data. *Regulatory toxicology and pharmacology* 66:241-247.
[31] Euling SY, Thompson CM, Chiu WA, Benson R. 2013. An approach for integrating toxicogenomic data in risk assessment: The dibutyl phthalate case study. *Toxicology and applied pharmacology* 271:324-335.

CHAPTER III

Evaluation and comparison of the relationship between NOEC and EC_{10}/EC_{20} values in chronic *Daphnia* toxicity testing

INTRODUCTION

Two measurement endpoints are commonly used to summarize toxicity in ecological risk assessment and ecotoxicology: the NOEC and the ECx. Each can be calculated from a variety of endpoints for daphnid species (e.g., survival, weight, total young produced). The hypothesis-based No Observed Effect Concentration (NOEC) is defined as the concentration at which there is no statistically significant difference from the control population for a measured endpoint [1]. The regression-based Effect Concentration (ECx) is the concentration at which there is an x% effect (reduction) at the measured endpoint; for instance, a reproductive EC_{50} signifies a 50% reduction in total young relative to the control [1]. These summary statistics are intended to provide useful toxicity information that is predictive of potential effects on the exposed population or ecosystem.

NOEC and ECx values are determined using procedures recommended by the Organisation for Economic Cooperation and Development (OECD), which establishes internationally developed test guidelines that are mutually accepted by member nations. Toxicity testing employs the NOEC or a combination of NOEC and ECx for multiple species across multiple industries and regulatory

entities. USEPA (United States Environmental Protection Agency) whole effluent testing guidelines rely on the NOEC and the IC_{25} (where I denotes inhibition) to assess effluent toxicity [2, 3]. Chronic Species Sensitivity Distributions (SSDs) use a NOEC and/or ECx for single-species toxicity to extrapolate concentrations that will be protective of the most species in an ecosystem [4, 5]. Risk assessors use NOECs (most frequently) or ECx to calculate a predicted no-effect concentration which is then used to predict safe levels for water quality or chemical toxicity in the environment [6, 7]. NOEC/ECx values indicating ecotoxicity are foundational to the European Union's Registration, Evaluation, Authorisation and Restriction of Chemicals (REACh) regulation [8].

Given that the choice of whether to use a NOEC or ECx is critically important to hazard determinations for chemical substances, close scrutiny of these statistics is not surprising. OECD sponsored a 1996 Workshop on Statistical Analysis of Aquatic Toxicity Data to review and compare data analysis options for ecotoxicity testing, including the NOEC and ECx [9]. The outcomes of this workshop included consensus on the need for NOEC replacement and directives to phase the NOEC out of OECD test guidelines, transitioning to an ECx measurement. The push towards replacement of the NOEC with ECx was supported by a series of tests focusing on inter-and intra-lab variability (ring test) using the OECD TG211 *Daphnia* reproductive toxicity guidelines, including a comparison of NOEC/ECx values across three compounds [10]. This data was useful in the modification of subsequent test guidelines to clarify a number of points of

concern, such as the use of test strategies to ensure sufficient power and refinement of test condition parameters to reduce variability. Based on the results of the workshop and ring test, it is now recommended to calculate and report ECx [11, 12] Furthermore, the philosophy extends to chronic algal (OECD 201 TG, [13]) and chronic fish (OECD 210 TG, [14]) testing.

Recent articles highlight division in the scientific community regarding retention and replacement of the NOEC [15-19]. Concerns raised about continuing use of the NOEC include the influence of experimental design (e.g., choice of concentrations, spacing of concentrations across the range) and technique issues (e.g. high control mortality) on the accuracy of NOEC calculation [20-22]. Peer-reviewed journals in medicine, psychology, and education disciplines have discouraged the use of hypothesis testing [23-25] and calls for a similar prohibition in ecotoxicology persist [18, 26, 27]. A major challenge is reaching consensus on which endpoint is the best, most viable replacement for the NOEC (e.g., EC_{10} or EC_{20}) [28, 29].

Without consensus on such vital points as the relationship between NOEC and ECx and the optimal ECx to replace the NOEC, a full analysis of the relationship of NOEC to ECx based on empirical data is essential. However, examples of such a comparison are difficult to find in published literature. A direct comparison of NOEC and ECx was included in the OECD ring test of *Daphnia* chronic toxicity, which only covered 3 compounds and was not meant as an exhaustive comparison of these endpoints [10]. An evaluation of the OECD 210 Fish Early-

Life Stage Test provided a limited comparison of NOEC/LOEC/ECx from the perspective of natural variability and its effect on statistical power [30]. A risk assessment of alcohol ethoxylates (AEs) showed a robust relationship between the AE chronic NOEC and EC_{10}s [31]. Lastly, a preliminary assessment by Isnard and Flammarion [29] of NOEC and ECx concluded that the most relevant replacement for NOEC was a low percent ECx.

This study addresses the need for a data-driven, literature-based evaluation of the suitability of EC_{10} or EC_{20} as a replacement for the NOEC, based on the association between NOEC and EC_{10}/EC_{20} values. The objectives of this study were to 1) compile a comprehensive chronic toxicity test dataset, then use it to compare NOEC to EC_{10} and EC_{20} values; 2) analyze the impact of endpoint choice on the relationship of NOEC to EC_{10} and EC_{20}; 3) evaluate the influence of test parameters on the relationship of NOEC to EC_{10} and EC_{20}; and 4) recommend alterations to future guidelines that utilize NOEC and ECx endpoints.

METHODS

Dataset Design and Construction

SIFT methodology outlined in Beasley et al. (in press) was employed to design, compile, and analyze the dataset used for this study.

SIFT Step 0: Dataset Construction

The master dataset was compiled from 200 *Daphnia* chronic toxicity test reports covering 36 years and including 90 unique chemicals. Tests were administered by 14 independent contract laboratories. A literature search was performed and ten additional peer-reviewed studies added to the master dataset for a total of 210 studies. Parameters and definitions are detailed in Table 1. Data was compiled in Microsoft Excel (2011).

SIFT Step 1: Relevance Criteria

Listed below are relevance criteria based on the defined purpose: evaluation of the relationship between NOEC and ECx values using chronic *Daphnia* tests. All studies failing these criteria were filtered out of the master dataset.

1) Tests conducted under OECD *Daphnia magna* reproduction test protocol TG211 or approved modification
2) Tests utilized *Daphnia magna* or *Ceriodaphnia dubia*
3) Reported NOEC value (or ascertained from raw data)
4) Effect data to the level of individual replicate

Based on these criteria, the original 210 test reports were filtered down to 156 viable test reports.

SIFT Step 2: Validity Criteria

Listed below are validity criteria based on OECD guidelines. All studies failing these criteria were filtered out of the master dataset:

1) Mortality of adult control populations (≤ 20%)

2) Mean number of offspring per surviving parent (≥ 60 by test end, *D. magna*)
3) Condition of organism (from the same culture, first instar, no ephippia, no males present)
4) Test duration of 21-23 days (*D. magna*) or 7 days (*C. dubia*)

Based on these criteria, the 156 studies from Step 1 were filtered down to 136 viable test reports.

SIFT Step 3: Acceptability Criteria

Listed below are acceptability criteria based on OECD guidelines. All studies failing these criteria were filtered out of the master dataset:

1) Actual concentrations reported
2) Minimum of 5 test concentrations plus control
3) Parameter information included or available: test type (semi-static, flow-through), test strategy (10x1, 4x10, 4x5, other), chemical class (neutral organic, polymer, anionic surfactant, cationic surfactant, nonionic surfactant, other), solvent used (water, other). In cases where chemical class was not included in the original test report, a search by CAS number in ECOSAR [32] was performed to determine class.
4) Sufficient dose response effect to calculate survival and ECx
5) No calculated hormetic effect

Based on these criteria the 136 studies from Step 2 were filtered down to 95 viable studies.

SIFT Step 4: Additional Criteria

Listed below are additional criteria based on OECD guidelines. All studies failing these criteria were filtered out of the master dataset:

1) Single compound only, no mixtures
2) At least two of the following four parameters included in the test report: temperature (in-range: 18-22°C), water hardness (in-range: ≥ 140 mg/L as $CaCO_3$), pH (in-range: 6-9), K_{ow}.

Based on these criteria the 95 studies from Step 3 were filtered down to a final dataset of 78 viable studies.

Table 1. Information collected from 210 Daphnia chronic toxicity test reports and peer-reviewed studies to populate a master dataset.

Test Parameter	Definition
Year	Year in which study was completed
Lab	Laboratory responsible for test administration
Chemical abstracts number (CAS)	Number and name assigned in CAS database; found in ECOSAR or CAS directly
Class	Chemical class as assigned by ECOSAR or expert judgment
Species	*Daphnia magna* or *Ceriodaphnia dubia*
Test type	Semi-static or flow-through configuration
Test strategy	Configuration of replicates
Test duration	Length of test in days
Solvent	Solvent in which test substance is dissolved prior to testing
Water	Ranges and means of temperature, hardness, and pH
Endpoints (as available)	Adult survival
	Total young produced
	Total young per surviving adult
	Total young per surviving adult per day
	Length (with standard deviation) in mm (total or per surviving adult)
	Dry/wet weight (with standard deviation) in g (total or per surviving adult)
	Days to first brood
	Number of broods
Effect data (as available, vary by study)	NOEC (No Observed Effect Concentration)
	LOEC (Lowest Observed Effect Concentration)
	EC_{10} (10% Effect Concentration) with confidence levels
	EC_{20} (20% Effect Concentration) with confidence levels
	Most sensitive endpoint at 10% effect
	Most sensitive endpoint at 20% effect

Identification and calculation of NOEC and ECx values

NOEC: NOEC values were taken directly from each test report unless not reported, in which case the reported LOEC was used to calculate NOEC by identifying the next test concentration below the reported LOEC.

ECx: Raw data was collected from each test report for the survival endpoint. EC_{10} and EC_{20} were calculated from this raw data in R [33] using the probit method. These EC_{10} and EC_{20} values were termed 'survival only' in further dataset analysis.

Additionally, raw data was collected from each test report for at least one reproductive endpoint. Any available raw data for additional reproductive endpoints was also collected. Using this raw data, EC_{10} and EC_{20} values were calculated in R using the Bruce-Versteeg model [34].

For each test report, the resulting calculated reproductive EC_{10} values only were compared to determine the lowest value, which was termed 'reproductive only' in further dataset analysis. This comparison was repeated for each test report to determine the reproductive only EC_{20} value.

To determine the most sensitive EC_{10} overall for each test report, the survival only EC_{10} was compared to all reproductive EC_{10} values. The lowest value in this comparison was termed the 'most sensitive EC_{10}'. This comparison was repeated for each test report to determine the most sensitive EC_{20} value.

Analysis of endpoint data

Distributions of the relative differences between NOEC:EC_{10} and NOEC:EC_{20} were tested for normality using a continuous fit and the Shapiro-Wilk W test. All were determined to be nonparametric. Wilcoxon Signed Rank analysis using the matched pairs function was completed for NOEC:ECx comparison within endpoints and across parameters. All parameters were evaluated for the most sensitive survival endpoint only, the most sensitive reproductive endpoint only, and the most sensitive endpoint overall. For all analyses the null hypothesis was that the NOEC and ECx were not significantly different. The null hypothesis was rejected when $p < 0.05$. Statistical analysis was completed using JMP 9.0.2 [35].

RESULTS

Results for the most sensitive endpoint are presented below. For the calculated ECx, the most sensitive endpoint was found to be mortality (51.7%) and reproduction (48.3%)(Figure 1). Survival-only and reproductive-only analyses are available in supplemental tables 1-6. Analysis of chemical class parameters is available in supplemental table 7.

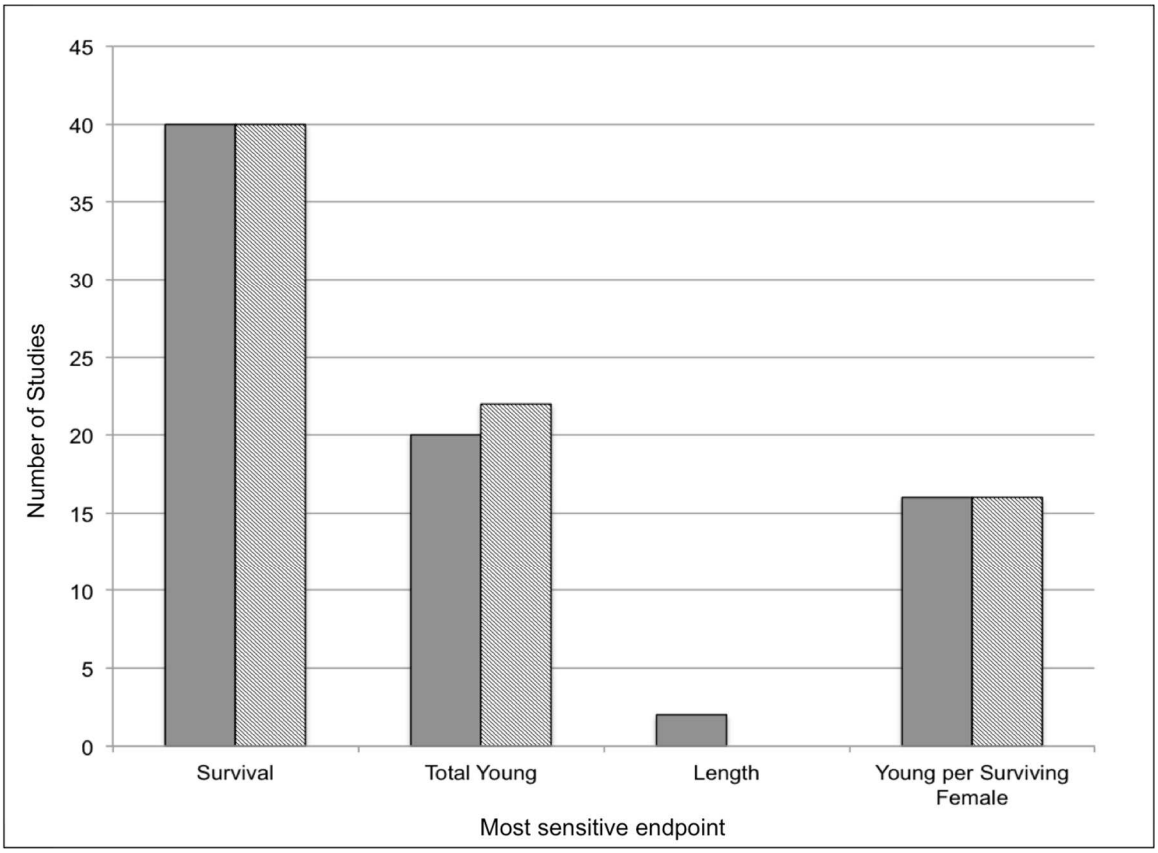

Figure 1. Most sensitive endpoints for calculated effect concentrations. Dark gray indicates a 10% effect concentration, while light gray is 20%. *n* = 78.

Test type

Analysis of the association between NOEC:EC_{10} and NOEC:EC_{20} showed that the NOEC was more strongly related to the EC_{10} than the EC_{20} ($|s|$ = 0.1812, n = 78 and $|s|$ = 0.0345, n = 78, respectively) and that NOEC was significantly different than EC_{20} (Table 2). Both test types showed a similar NOEC:EC_{10} relationship (semi-static, $|s|$ = 0.3531, n = 36 and flow-through, $|s|$ = 0.3644, n = 42, respectively). The NOEC/EC_{20} relationship was weak for semi-static and flow-through ($|s|$ = 0.2193, n = 36 and $|s|$ = 0.0787, respectively).

Table 2. Analysis of NOEC:EC_{10} and NOEC:EC_{20} relationships for the Test Type parameter, where |*s*| = rank sum *p*-value and *n* = 78. Italics denote significant difference at $\alpha = 0.05$.

Test Type		|s|, NOEC:EC10	n	|s|, NOEC:EC20	n
Most sensitive	**All test types**	**0.1812**	**78**	***0.0345***	**78**
Most sensitive	Semi	0.3531	36	0.2193	36
Most sensitive	Flow	0.3644	42	0.0787	42

Species

The NOEC:EC_{10} was more strongly related than NOEC:EC_{20} (Table 3). This pattern was similar when analyzing *C. dubia* (|*s*| = 0.7606, *n* = 24 and |*s*| = 0.0129, *n* = 24, respectively). *D. magna* patterns were reversed, with NOEC:EC_{20} more closely associated than NOEC:EC_{10} (EC_{10} |*s*| = 0.1485, *n* = 54 and EC_{20} |*s*| = 0.3778, *n* = 54).

Table 3. Analysis of NOEC:EC_{10} and NOEC:EC_{20} relationships for the Species parameter, where |s|= rank sum p-value and n=78. Italics denote significant difference at $\alpha=0.05$.

Species		|s|, NOEC:EC10	n	|s|, NOEC:EC20	n
Most sensitive	**All species**	**0.1812**	**78**	***0.0345***	**78**
Most sensitive	*D. magna*	0.1485	54	0.3778	54
Most sensitive	*C. dubia*	0.7606	24	*0.0129*	24

Strategy

Analysis of the strategy parameter showed NOEC:EC_{10} was more strongly related than NOEC:EC_{20} (Table 4). This pattern repeated for the 10x1 experimental design strategy (EC_{10} |*s*| = 0.9253, *n* = 47 and EC_{20} |*s*| = 0.0029,

n = 47) as well as for the 4x10 strategy (EC_{10} |*s*| = 0.8311, *n* = 11 and EC_{20} |*s*| = 0.1016, *n* = 11). This pattern was reversed for the 4x5 and Other strategies. Both EC_{10} and EC_{20} results as compared to NOEC for the 4x5 were significantly different (EC_{10} |*s*| = 0.0093, *n* = 12 and EC_{20} |*s*| = 0.0425, *n* = 12).

Table 4. Analysis of NOEC:EC_{10} and NOEC:EC_{20} relationships for the Strategy parameter, where |*s*| = rank sum *p*-value and *n* = 78. Italics denote significant difference at α = 0.05.

Strategy		|s|, NOEC:EC10	n	|s|, NOEC:EC20	n
Most sensitive	**All strategies**	**0.1812**	**78**	***0.0345***	**78**
Most sensitive	10x1	0.9253	47	*0.0029*	47
Most sensitive	4x5	*0.0093*	12	*0.0425*	12
Most sensitive	4x10	0.8311	11	0.1016	11
Most sensitive	Other	0.25	8	0.8438	8

pH

Evaluation of pH showed that NOEC and EC_{20} were more strongly related (|*s*| = 0.5672, *n* = 62) than EC_{10} (|*s*| = 0.0288, *n* = 62)(Table 5). Of the 62 test reports documented, 59 were in range. Data from these in-range tests showed a stronger relationship than the overall dataset for NOEC:EC_{20} (|*s*| = 0.8756, *n* = 59) and a significant difference between the NOEC and the EC_{10} (|*s*| = 0.0081, *n* = 59).

Table 5. Analysis of NOEC:EC_{10} and NOEC:EC_{20} relationships for the pH parameter, where |s|= rank sum p-value and n=62. Italics denote significant difference at α=0.05.

pH		|s|, NOEC:EC10	n	|s|, NOEC:EC20	n
Most sensitive	**All pH values**	***0.0288***	**62**	**0.5672**	**62**
Most sensitive	Range	*0.0081*	59	0.8756	59
Most sensitive	High	0.25	3	0.25	3

Hardness

Analysis of hardness showed that NOEC:EC_{20} was more strongly related than NOEC:EC_{10} (EC_{20} |*s*| = 0.7053, *n* = 60 and EC_{10} |*s*| = 0.0216, *n* = 60, respectively)(Table 6). For both low and in-range hardness values, the NOEC was more strongly related to the EC_{20} (low |*s*| = 0.6322, *n* = 16 and range |*s*| = 0.427, *n* = 44) than the EC_{10} (low |*s*| = 0.1165, *n* = 16 and range |*s*| = 0.0979, *n* = 44).

Table 6. Analysis of NOEC:EC_{10} and NOEC:EC_{20} relationships for the Hardness parameter, where |*s*| = rank sum *p*-value and *n* = 60. Italics denote significant difference at *α* = 0.05.

Hardness		|s|, NOEC:EC10	n	|s|, NOEC:EC20	n
Most sensitive	**All hardness values**	***0.0216***	**60**	**0.7053**	**60**
Most sensitive	Low	0.1165	16	0.6322	16
Most sensitive	Range	0.0979	44	0.427	44

Temperature

Temperature analysis showed a significant difference between NOEC and EC_{20} (|*s*| = 0.0345, *n* = 78) and no difference between NOEC and EC_{10} (|*s*| = 0.1812, *n* = 78) (Table 7). This pattern was also observed when analyzing in-range temperature (EC_{10} |*s*| = 0.1718, *n* = 73; EC_{20} |*s*| = 0.0548, *n* = 73, respectively)

and high temperature (EC_{10} |*s*| = 0.3125, *n* = 5 and EC_{20} |*s*| = 0.1875, respectively).

Table 7. Analysis of NOEC:EC_{10} and NOEC:EC_{20} relationships for the Temperature parameter, where |*s*| = rank sum *p*-value and *n* =78. Italics denote significant difference at α = 0.05.

Temperature		|s|, NOEC:EC10	n	|s|, NOEC:EC20	n
Most sensitive	**All temperature values**	**0.1812**	**78**	***0.0345***	**78**
Most sensitive	Range	0.1718	73	0.0548	73
Most sensitive	High	0.3125	5	0.1875	5

DISCUSSION

If the ECx is to replace the NOEC, an important task is to determine the most accurate and practical ECx value: a concentration that is high enough so that effects can be reliably calculated from dose-response data, while still low enough to protect sensitive species and therefore the exposed population. Bruce and Versteeg (1992) estimated a 20% effect concentration as an environmentally relevant concentration that would minimize adverse effects on a population as compared to natural variability. Others have set this minimal effect threshold at 25% [36, 37]. Still others contend that an effect concentration of 10% is a more realistic estimate of relative hazard [12, 28, 38]. Traditionally, some researchers have treated the EC20 as a surrogate to the NOEC, with the justification that no statistically significant difference to the control still results in a measurable effect on the order of 10-20%[30, 34, 39]. In other words, “no effect” is a statistical misnomer. Probing this question with robust ecological process models could

inform to what degree X values of 10 or 20 would result in altered population community structure. If 20% adverse response relative to the control condition did not result in altered population and community structure, one could conclude that the EC_{20} provides an appropriately conservative analogue of a NOEC value. Some risk assessors continue to use a 20% effect as an appropriate NOEC surrogate [40-42]. Ecotoxicology testing methods have gradually shifted towards equating NOEC values with a 10% effect [43]. This shift is influenced by the 1994 OECD *Daphnia* chronic ring test, in which the tested EC_{20}s were higher than the NOEC and the tested EC_{10}s were scattered around the NOEC [10]. Environment Canada and Australia/New Zealand have already moved to replace NOEC with EC_{10} in aquatic toxicology and water safety standards [31, 44]. The present study does not attempt to argue whether the NOEC should be replaced or whether the ECx is a useful replacement. Clearly the shift towards an ECx replacement of NOEC is underway although the best NOEC surrogate has yet to be established. The present study applied a meta-analysis approach focused on strength and significance of association to assess suitability of the EC_{10} or EC_{20} as a potential replacement for the NOEC. We found that overall the NOEC was more strongly related to the EC_{10} than the EC_{20}.

The choice of chronic *Daphnia* toxicity data was advantageous for this study due to the large volume of information available from a variety of independent sources and over a long period of time. *Daphnia* test data allowed for a

straightforward analysis as it generally follows OECD standard test guidelines, which are similar to other invertebrate chronic toxicity protocols and have remained relatively static over time. These advantages allow confidence in the depth and breadth of the dataset, its relevance to a wide range of applications, and the ability to extrapolate evident trends to other contexts in which NOEC:ECx is useful.

Although the *Daphnia* chronic test is designed to use reproductive endpoints to explain effect, these endpoints do not always return the most sensitive and therefore the most conservative value. In the present study, mortality endpoints actually comprised 51% of the most sensitive calculated EC_{10}s (Figure 1). A variety of growth and reproductive endpoints (e.g. length, weight, time to first brood, total young per surviving adult per day) were used to calculate NOEC or ECx in the original test reports; of these, the required 'total number of offspring per surviving parent' was not always the most sensitive. The variety of endpoints returning the most sensitive value in ECx calculations speaks to the potential usefulness of recording several measurement endpoints to give a complete picture of chronic toxicity.

Current OECD test guidelines recommend flow-through test design and allow the use of traditional semi-static designs with certain restrictions [12]. Flow-through delivery is commonly thought to ensure consistent dose distribution but return

less sensitive toxicity measurement, while semi-static is considered more sensitive but with less control over test concentrations [10, 12, 45]. In the present study, both test types were similarly represented (semi-static *n* = 36, flow-through *n* = 42) (Table 2). Given the relative advantages of each, it is not surprising that this dataset showed test type choice did not impact the relative strength of NOEC:EC_{10} relationship as compared to NOEC:EC_{20}.

The *Daphnia* reproduction chronic toxicity test [12] protocol is designed for the use of *D. magna*, with the option to use other daphnid species if justified. *Ceriodaphnia dubia* have been popular in the past because of shorter life span and high sensitivity [46], which permits testing similar to OECD 211 but with a 7-day duration. The NOEC/EC_{10} was strongly related in the dataset as a whole and very strongly when analyzing for *C. dubia* alone (Table 3). Apparent impact of species choice to the strength of NOEC:$EC_{10,}$ relationship may be attributed to overall test quality, as *C. dubia* use was concentrated in a group of strictly controlled tests. A larger dataset might provide clarity to species parameter analysis.

Although the choice of test type and species appeared to have little impact on the NOEC:EC_{10} relationship, hidden in that data is the impact of an interrelated parameter: test strategy. Ambiguous test strategy recommendations in previous OECD guidelines were addressed in the updated 2012 TG211, covering issues

surrounding the accurate measurement of parental mortality and number of offspring per parent [12]. Now semi-static tests require the use of a 10x1 strategy (10 replicates of 1 adult organism). Flow-through tests are allowed leeway in choice of strategy, although a 4x10 strategy is recommended. In the present study, analysis of the most sensitive endpoint showed an overall pattern of stronger NOEC:EC_{10} relationship, including a more strongly related NOEC:EC_{10} when 10x1 and 4x10 strategies were used (Table 4). Tests conducted under the previous set of guidelines reflect the variability addressed in the TG211 revisions, where the '4x5' and 'Other' strategies show a stronger relationship between NOEC and EC_{20}.

For the *Daphnia* chronic test to be valid or acceptable, the OECD guideline requires a few key test conditions to lie within specified ranges; other conditions are flexible or unrestricted. pH was an example of a flexible guideline recommendation with a broad range of acceptable values. Guidelines state that pH within test vessels should be between 6-9, although it is noted that the pH should not vary more than 1.5 pH units throughout a test. In the present study, test reports did not address the variability of pH throughout the test and only reported initial or final pH values, as is common in routine toxicity testing. Sixteen test reports did not include pH information at all and three test records reported a pH outside range. pH has been shown to influence *Daphnia* fitness and reproduction, including detrimental effects to respiration between pH 6-7 [47,

48] and viable egg production at pH 9 [49]. This would suggest that low or high pH could affect reproductive rate [50, 51] and therefore ECx calculations. The broad, flexible pH guideline introduced variability that likely contributed to a stronger NOEC:EC_{20} relationship and significant difference between NOEC and EC_{10} (Table 5). Based on the influence of pH on *Daphnia* combined with analysis in the present study, pH guidelines should be defined and restricted beyond the current suggested 6-9 range, including reporting of pH values.

An even less restricted parameter is hardness. Although water hardness is important in *Daphnia* fitness, guidelines only suggest that for *D. magna* hardness should be at or above 140 mg/L $CaCO_3$. 23% of test reports did not report hardness, and of those that did, in-range values extended from 140 to 450 mg/L. Lewis and Maki [52] found *D. magna* produced 65% more young when reared in 350 mg/L $CaCO_3$ as compared to 50 mg/L CaCO3, and Paulauskis and Winner [53] found that increasing hardness from 50 to 200 mg/L CaCO3 significantly affected brood size. Hardness is well known to affect the bioaccumulation and toxicity of metals and other compounds [51, 54]. The present study showed that the NOEC:EC_{20} was more strongly related for all endpoints, reinforcing the influence of hardness on ECx (Table 6). Like pH, this hardness data underscores the need for further analysis and clarification of rationales for guideline-driven suggestions on water quality measurements, water sources, and experimental design.

As a counterpoint to analyses of pH and hardness parameters, the temperature parameter guidelines are well defined within OECD guidelines. A broad range of testing temperature (18-22°C) is acceptable; unlike pH and hardness, temperature must be measured daily given the direct relationship between temperature and developmental rate in aquatic life [55]. In addition, daily temperatures should remain within a 2°C range (e.g., 20-22°C). In the present study, all test reports included temperature information. NOEC:ECx relational patterns for temperature closely followed those of the overall dataset, where NOEC was more strongly related to EC_{10} as compared to EC_{20} (Table 7). Temperature affects rate and efficacy of reproduction in *Daphnia* [56]; tightly restricting temperature guidelines likely minimized this parameter's impact on ECx.

CONCLUSION

An objective, methodical evaluation of the association between NOEC and EC_{10} and NOEC and EC_{20} was accomplished. Endpoint and parameter analysis revealed that each impacts the strength and significance of the relationship between NOEC and EC_{10}.

Evaluations of two key water quality parameters, pH and hardness, illustrate the importance of constraints on test conditions for optimal EC_x calculation. Analysis of pH and hardness parameters showed NOEC:EC_{20} was more strongly related than NOEC:EC_{10}, indicating that variability in test implementation may have

compounded variability in EC_x calculations. However, when parameters were more restricted and required (e.g. temperature), NOEC:EC_{10} was consistently more strongly related than NOEC:EC_{20}. Evaluation of the strategy parameter suggests that recent revisions to test guidelines (e.g., requirement to use 10x1 or 4x10 configurations) were effective in minimizing such variability. Daphnids may optimally develop under a more stringent set of water quality parameters and outside this range daphnids may have suitable overall viability, but increasingly variable response profiles (e.g., pH and hardness).

Based on this analysis of chronic toxicity test data, the EC_{10} is a more appropriate analogue for the NOEC than the EC_{20}. We recommend reporting of pH and hardness values due to their unique impact on EC_x calculations. We recommend refinement of the acceptable ranges for pH and hardness in order to minimize error in EC_x calculation.

REFERENCES

[1] OECD. 2006. Current Approaches in the Statistical Analysis of Ecotoxicology Data: A Guidance to Application. Series on Testing and Assessment, No. 54. Environmental Health and Safety Publications, Series on testing and assessment. *OECD Publishing*.

[2] Weber CI, Peltier W, Norberg-King T, Horning W, Kessler F. 1989. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms. Environmental Protection Agency, Cincinnati, OH (USA). Environmental Monitoring Systems Lab.

[3] USEPA USEPA. 2002. Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Water to Freshwater Organisms In Office of Water (4303), ed, Fourth ed. Vol EPA-821-R-02-013

[4] Versteeg DJ, Belanger SE, Carr GJ. 1999. Understanding single-species and model ecosystem sensitivity: Data-based comparison. *Environmental toxicology and chemistry* 18:1329-1346.

[5] Newman MC, Ownby DR, Mezin LC, Powell DC, Christensen TR, Lerberg SB, Anderson BA. 2000. Applying species-sensitivity distributions in ecological risk assessment: Assumptions of distribution type and sufficient numbers of species. *Environmental toxicology and chemistry* 19:508-515.

[6] Van der Hoeven N. 2004. Current issues in statistics and models for ecotoxicological risk assessment. *Acta biotheoretica* 52:201-217.

[7] De Laender F, Van Sprang P, Janssen CR. 2013. A re-evaluation of fifteen years of European risk assessment using effect models. *Environmental toxicology and chemistry* 32:594-601.

[8] European Parliament, The Council of the European Union. 2006. Regulation (EC) no. 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the registration, evaluation, authorization and restriction of chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) no. 793/93 and Commission Regulation (EC) no. 1488/94, as well as Council Directive 76/769/EEC and commission Directives 91/155/EEC, 93/67/EEC, 93/105/CE and 2000/21/EC. *Official journal of the European Union* 396.

[9] OECD. 1998. *Report of the OECD Workshop on Statistical Analysis of Aquatic Toxicity Data*. OECD Publishing.

[10] OECD. 1997. *Report of the Final Ring Test of the Daphnia magna Reproduction Test*. OECD Publishing.

[11] OECD. 2013. *Test No. 236: Fish Embryo Acute Toxicity (FET) Test*. OECD Publishing.

[12] OECD. 2012. *Test No. 211: Daphnia magna Reproduction Test*. OECD Publishing.

[13] OECD. 2011. *Test No. 201: Freshwater Alga and Cyanobacteria, Growth Inhibition Test*. OECD Publishing.

[14] OECD. 1992. *Test No. 210: Fish, Early-Life Stage Toxicity Test*. OECD Publishing.
[15] Warne MSJ, van Dam R. 2008. NOEC and LOEC data should no longer be generated or used. *Australasian journal of ecotoxicology* 14:1.
[16] Green JW, Springer TA, Staveley JP. 2013. The drive to ban the NOEC/LOEC in favor of ECx is misguided and misinformed. *Integrated environmental assessment and management* 9:12-16.
[17] Van der Vliet L, Taylor LN, Scroggins R. 2012. NOEC: Notable oversight of enlightened Canadians: A response to van Dam et al.(2012). *Integrated environmental assessment and management* 8:397-398.
[18] Jager T. 2012. Bad habits die hard: The NOEC's persistence reflects poorly on ecotoxicology. *Environmental toxicology and chemistry* 31:228-229.
[19] Landis WG, Chapman PM. 2011. Well past time to stop using NOELs and LOELs. *Integrated environmental assessment and management* 7:vi-viii.
[20] Chapman PM, Caldwell RS, Chapman PF. 1996. A warning: NOECs are inappropriate for regulatory use. *Environmental toxicology and chemistry* 15:77-79.
[21] Fox DR, Billoir E, Charles S, Delignette-Muller ML, Lopes C. 2012. What to do with NOECS/NOELS—prohibition or innovation? *Integrated environmental assessment and management* 8:764-766.
[22] Kooijman S. 1996. An alternative for NOEC exists, but the standard model has to be abandoned first. *Oikos* 75:310-316.
[23] Altman DG, Gore SM, Gardner MJ, Pocock SJ. 1983. Statistical guidelines for contributors to medical journals. *British medical journal (Clinical research ed)* 286:1489.
[24] Wilkinson L. 1999. Statistical methods in psychology journals: Guidelines and explanations. *American psychologist* 54:594.
[25] Thompson B. 1996. Research news and comment: AERA editorial policies regarding statistical significance testing: three suggested reforms. *Educational researcher* 25:26-30.
[26] van Dam RA, Harford AJ, Warne MSJ. 2012. Time to get off the fence: The need for definitive international guidance on statistical analysis of ecotoxicity data. *Integrated environmental assessment and management* 8:242-245.
[27] Bleaney A. 2012. Letter to the Editor. *International journal of environmental studies* 69:217-219.
[28] Belanger S, Dorn P, Toy R, Boeije G, Marshall S, Wind T, Van Compernolle R, Zeller D. 2006. Aquatic risk assessment of alcohol ethoxylates in North America and Europe. *Ecotoxicology and environmental safety* 64:85-99.
[29] Isnard P, Flammarion P, Roman G, Babut M, Bastien P, Bintein S, Essermeant L, Férard J, Gallotti-Schmitt S, Saouter E. 2001. Statistical analysis of regulatory ecotoxicity tests. *Chemosphere* 45:659-669.
[30] Oris JT, Belanger SE, Bailer AJ. 2012. Baseline characteristics and statistical implications for the OECD 210 fish early-life stage chronic toxicity test. *Environmental toxicology and chemistry* 31:370-376.

[31] Belanger S, Dorn P, Burridge L, Haya K, Niimi A. 2004. Chronic aquatic toxicity of alcohol ethoxylate (AE) surfactants under Canadian exposure conditions. 31st Annual Aquatic Toxicity Workshop, Charlottetown, Prince Edward Island, Canadian Technical Report of Fisheries and Aquatic Sciences.
[32] U.S. Environmental Protection Agency. 2012. ECOTOX User Guide: ECOTOXicology Database System. Version 4.0. Washington, DC.
[33] R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
[34] Bruce RD, Versteeg DJ. 1992. A statistical procedure for modeling continuous toxicity data. *Environmental toxicology and chemistry* 11:1485-1494.
[35] JMP 9.0.2. SAS Institute, Cary, North Carolina, USA.
[36] Norberg-King TJ. 1993. A linear interpolation method for sublethal toxicity: The inhibition concentration (ICp) approach. *National effluent toxicity assessment center technical report* 39:3-93.
[37] Barnthouse LW, Suter GW, Rosen AE, Beauchamp JJ. 1987. Estimating responses of fish populations to toxic contaminants. *Environmental toxicology and chemistry* 6:811-824.
[38] DeForest DK, Adams WJ, Chapman PM. 2008. What is an appropriate level of protection? An example considering selenium exposures by aquatic birds. *Integrated environmental assessment and management* 4:513-515.
[39] Grist EP, Wells NC, Whitehouse P, Brighty G, Crane M. 2003. Estimating the effects of 17α-ethinylestradiol on populations of the fathead minnow Pimephales promelas: Are conventional toxicological endpoints adequate? *Environmental science & technology* 37:1609-1616.
[40] Versteeg D, Rawlings J. 2003. Bioconcentration and toxicity of dodecylbenzene sulfonate (C12LAS) to aquatic organisms exposed in experimental streams. *Archives of environmental contamination and toxicology* 44:0237-0246.
[41] Barnthouse LW, Munns Jr WR, Sorensen MT. 2007. *Population-level ecological risk assessment*. CRC Press.
[42] Dyer S, Lauth J, Morrall S, Herzog R, Cherry D. 1997. Development of a chronic toxicity structure–activity relationship for alkyl sulfates. *Environmental toxicology and water quality* 12:295-303.
[43] Guilhermino Lc, Sobral OÃ, Chastinet C, Ribeiro R, Gonzalves F, Silva M, Soares AM. 1999. A *Daphnia magna* first-brood chronic test: an alternative to the conventional 21-day chronic bioassay? *Ecotoxicology and environmental safety* 42:67-74.
[44] Warne MSJ, Batley G, Braga O, Chapman J, Fox D, Hickey C, Stauber J, Van Dam R. 2014. Revisions to the derivation of the Australian and New Zealand guidelines for toxicants in fresh and marine waters. *Environmental science and pollution research* 21:51-60.
[45] Diamantino TC, Ribeiro R, Goncalves F, Soares AM. 1997. METIER (modular ecotoxicity tests incorporating ecological relevance) for difficult

substances. 4. Test chamber for cladocerans in flow-through conditions. *Environmental toxicology and chemistry* 16:1234-1238.
[46] Versteeg D, Stalmans M, Dyer S, Janssen C. 1997. Ceriodaphnia and Daphnia: A comparison of their sensitivity to xenobiotics and utility as a test species. *Chemosphere* 34:869-892.
[47] Alibone M, Fair P. 1981. The effects of low pH on the respiration of Daphnia magna Straus. *Hydrobiologia* 85:185-188.
[48] Havas M, Hutchinson TC, Likens GE. 1984. Effect of low pH on sodium regulation in two species of Daphnia. *Canadian journal of zoology* 62:1965-1970.
[49] Vijverberg J, Kalf DF, Boersma M. 1996. Decrease in Daphnia egg viability at elevated pH. *Limnology and oceanography* 41:789-794.
[50] Locke A, Sprules WG. 2000. Effects of acidic pH and phytoplankton on survival and condition of Bosmina longirostris and Daphnia pulex. *Hydrobiologia* 437:187-196.
[51] Belanger SE, Cherry DS. 1990. Interacting effects of pH acclimation, pH, and heavy metals on acute and chronic toxicity to Ceriodaphnia dubia (Cladocera). *Journal of crustacean biology* 10(2):225-235.
[52] Lewis M, Maki A. 1981. Effects of water hardness and diet on productivity of Daphnia magna Straus. in laboratory culture. *Hydrobiologia* 85:175-179.
[53] Paulauskis JD, Winner RW. 1988. Effects of water hardness and humic acid on zinc toxicity to Daphnia magna Straus. *Aquatic toxicology* 12:273-290.
[54] Long KE, Van Genderen EJ, Klaine SJ. 2004. The effects of low hardness and pH on copper toxicity to Daphnia magna. *Environmental toxicology and chemistry* 23:72-75.
[55] Beasley A, Elrod-Erickson M, Otter RR. 2012. Consistency of morphological endpoints used to assess developmental timing in zebrafish (Danio rerio) across a temperature gradient. *Reproductive toxicology* 34:561-567.
[56] Stephenson R, Watts S. 1984. Chronic toxicity tests with Daphnia magna: the effects of different food and temperature regimes on survival, reproduction and growth. *Environmental pollution series A, ecological and biological* 36:95-107.

APPENDICES

APPENDIX A: Tables

Supplemental Files

Table 1. Analysis of NOEC:EC_{10} and NOEC:EC_{20} relationships for the Test Type parameter, where |*s*| = rank sum *p*-value and *n* = 78. Italics denote significant difference at α = 0.05.

Test Type		NOEC/10%	n	NOEC/20%	n
Survival alone	**All test types**	**0.0808**	**78**	**0.0003**	**78**
Survival alone	Semi	0.4873	36	0.048	36
Survival alone	Flow	0.1034	42	*0.0017*	42
Repro alone	**All test types**	**0.0013**	**78**	**0.0001**	**78**
Repro alone	Semi	*0.0205*	36	*0.0001*	36
Repro alone	Flow	*0.0264*	42	*0.0001*	42

Table 2. Analysis of NOEC:EC_{10} and NOEC:EC_{20} relationships for the Species parameter, where |s|= rank sum p-value and n=78. Italics denote significant difference at α=0.05.

Species		NOEC/10%	n	NOEC/20%	n
Survival alone	**All species**	**0.0808**	**78**	**0.0003**	**78**
Survival alone	*D. magna*	0.5152	54	0.1802	54
Survival alone	*C. dubia*	*0.0001*	24	*0.0001*	24
Repro alone	**All species**	**0.0013**	**78**	**0.0001**	**78**
Repro alone	*D. magna*	*0.0003*	54	*0.0001*	54
Repro alone	*C. dubia*	0.7396	24	*0.0006*	24

Table 3. Analysis of NOEC:EC_{10} and NOEC:EC_{20} relationships for the Strategy parameter, where |s|= rank sum p-value and n=78. Italics denote significant difference at α=0.05.

Strategy		**NOEC/10%**	**n**	**NOEC/20%**	**n**
Survival alone	**All strategies**	**0.0808**	**78**	***0.0003***	**78**
Survival alone	10x1	*0.0003*	47	*0.0001*	47
Survival alone	4x5	*0.0034*	12	*0.0425*	12
Survival alone	4x10	0.9658	11	0.1016	11
Survival alone	Other	0.3125	8	0.9453	8
Repro alone	**All strategies**	***0.0013***	**78**	***0.0001***	**78**
Repro alone	10x1	*0.0117*	47	*0.0001*	47
Repro alone	4x5	0.2661	12	0.0771	12
Repro alone	4x10	0.1016	11	*0.0098*	11
Repro alone	Other	0.4609	8	0.1953	8

Table 4. Analysis of NOEC/EC_{10} and NOEC/EC_{20} relationships for the pH parameter, where |s|= rank sum p-value and n=62. Italics denote significant difference at α=0.05.

pH					
Survival alone	**All pH values**	**0.1125**	**62**	***0.0001***	**62**
Survival alone	Range	0.1162	59	0.5085	59
Survival alone	High	0.25	3	0.25	3
Repro alone	**All pH values**	***0.0016***	**62**	***0.0005***	**62**
Repro alone	Range	*0.0081*	59	*0.0001*	59
Repro alone	High	0.25	3	0.25	3

Table 5. Analysis of NOEC:EC_{10} and NOEC:EC_{20} relationships for the Hardness parameter, where |s|= rank sum p-value and n=60. Italics denote significant difference at α=0.05.

Hardness					
Survival alone	**All hardness values**	**0.4529**	**60**	**0.1899**	**60**
Survival alone	Low	0.2522	16	0.8209	16
Survival alone	Range	0.7567	44	0.1568	44
Repro alone	**All hardness values**	***0.0045***	**60**	***0.0001***	**60**
Repro alone	Low	0.1046	16	*0.0155*	16
Repro alone	Range	*0.0248*	44	*0.0001*	44

Table 6. Analysis of NOEC:EC_{10} and NOEC:EC_{20} relationships for the Temperature parameter, where |s|= rank sum p-value and n=78. Italics denote significant difference at α=0.05.

Temperature		**NOEC/10%**	**n**	**NOEC/20%**	**n**
Survival alone	**All temperature values**	**0.0808**	**78**	***0.0003***	**78**
Survival alone	Range	0.1108	73	*0.0008*	73
Survival alone	High	0.3125	5	0.1875	5
Repro alone	**All temperature values**	***0.0013***	**78**	***0.0001***	**78**
Repro alone	Range	*0.003*	73	*0.0001*	73
Repro alone	High	0.0625	5	0.0625	5

Table 7. Analysis of NOEC:EC_{10} and NOEC:EC_{20} relationships for the Class parameter, where |s|= rank sum p-value and n=78. Italics denote significant difference at α=0.05.

Class		NOEC/10%	n	NOEC/20%	n
Survival alone	**All classes**	**0.0808**	**78**	***0.0003***	**78**
Survival alone	Nonionic Surfactant	0.5625	6	0.2188	6
Survival alone	Anionic Surfactant	*0.0059*	26	*0.0001*	26
Survival alone	Cationic Surfactant	0.4697	12	0.3394	12
Survival alone	Polymer	0.2188	6	0.1563	6
Survival alone	Neutral Organic	0.9097	12	0.5186	12
Survival alone	Other	0.1754	16	0.7436	16
Repro alone	**All classes**	***0.0013***	**78**	***0.0001***	**78**
Repro alone	Nonionic Surfactant	0.8438	6	0.0625	6
Repro alone	Anionic Surfactant	0.1403	26	*0.0001*	26
Repro alone	Cationic Surfactant	*0.0269*	12	*0.0122*	12
Repro alone	Polymer	0.5625	6	0.0625	6
Repro alone	Neutral Organic	0.2661	12	0.0923	12
Repro alone	Other	0.0739	16	*0.0131*	16

CHAPTER IV

Influence of chemical class, endpoint selection and application factor parameters on thresholds of toxic concern in ecotoxicological hazard assessment

INTRODUCTION

The threshold of toxic concern (TTC) is an alternative tool for risk assessment that has been well established in the food safety and pharmaceutical sectors, where the TTC has been used for decades to clear food additives and pharmaceutical impurities [1-4]. This tool is built on the accepted concept that 'safe' chemical concentrations exist at which there is no risk to human health [2] Furthermore, the threshold idea can extend to predicting safe concentrations for unknown or untested chemicals based on similar structures or modes of action [5]. The TTC has been expanded to include screening for multiple endpoints, including carcinogenicity and genotoxicity, in multiple routes of exposure [6,7].

The TTC concept has been tentatively explored in the context of ecotoxicological risk assessment. In general, a set of ecotoxicity data endpoints is compiled from which a lowest or 95th percentile value (HC_5 or hazard concentration, 5%) is identified as a 'safe' concentration; specific thresholds could be calculated based on criteria such as structure, chemical class, or mode of action. Recent applications include proposed TTC-like values for endocrine-active substances [8] and a similar study by de Wolf and colleagues of organic chemicals [9].

Gutsell et al. (2015) similarly analyzed a set of consumer product chemicals for a threshold value, specifically noting cationically charged chemicals [10]. Thresholds were similar to those shown by de Wolf [9] excepting cationics, which were significantly lower than thresholds for other classes.

Advantages to the use of an ecotoxicological TTC (ecoTTC) include the ability to rapidly screen chemicals produced at low volumes or those for which little data exists. Difficult chemicals, such as those with no available quantitative structure-activity relationship (QSAR) model, would benefit from the addition of an ecoTTC to the toxicity profile. Elimination of the time, expense, and animal use associated with a traditional toxicity testing program make the TTC approach especially appealing in light of REACH registration requirements.

In this study, a preliminary exploration of the ecoTTC concept was undertaken to better understand the scope and limitations of implementing such a concept. The objectives were to 1) construct a metadataset using public and private-sector toxicity data that would allow a broad, deep analysis; and 2) perform preliminary evaluations of the impact of endpoint selection, application factor, and chemical class parameters on ecoTTC.

METHODS

Dataset construction

The SIFT method (Beasley et al., in press) was used to select acute and chronic test data based on species and test guideline to include a variety of test chemicals and taxa. Taxa were defined as algal, invertebrate, or vertebrate (fish). Chemical Abstract Service (CAS) registry numbers for each chemical were assigned based on ECOSAR [11] designation or expert judgment. Chemical classes were assigned based on ECOSAR designation or expert judgment. ECOSAR acute toxicity modeling endpoints were collected for each CAS when available. Master dataset architecture was based on unique CAS numbers as previously noted.

Calculations

All endpoints were normalized to mg/L for ease of subsequent calculation. In cases where only one test result existed for a CAS or chemical, Geometric mean was calculated first for each common endpoint for a species within a study (e.g., CAS 1234, Study X, Daphnia magna, NOEC). Geometric mean was then calculated across acute/chronic for each taxon per study per CAS (e.g., CAS 1234, Study X, invertebrates, acute endpoints) so as to provide one acute or chronic endpoint per taxon per study.

Geometric means were recorded such that application factors could be correctly calculated based on numbers of means available (e.g., CAS 1234, 3 acute

invertebrate geometric means, 1 chronic algal mean). Most sensitive endpoint was selected for each CAS from available acute and chronic endpoints across all three taxa.

Application factors were utilized based on United States Environmental Protection Agency (USEPA) and Organisation for Economic Cooperation and Development (OECD) [12, 13]. In cases where guidelines referenced 'most sensitive' species in selection of appropriate application factors, ECOSAR modeling data provided supporting information. PNECs (ecological predicted no effect concentrations) were calculated for each CAS using most sensitive endpoint, ECOSAR modeling information when needed, and appropriate application factor for both North American and European use. Lognormal distributions of PNECs and resulting hazard concentrations (HC5) were calculated in R [14].

RESULTS

The preliminary ecoTTC metadataset was created and analyzed using the process diagram shown in Figure 1.

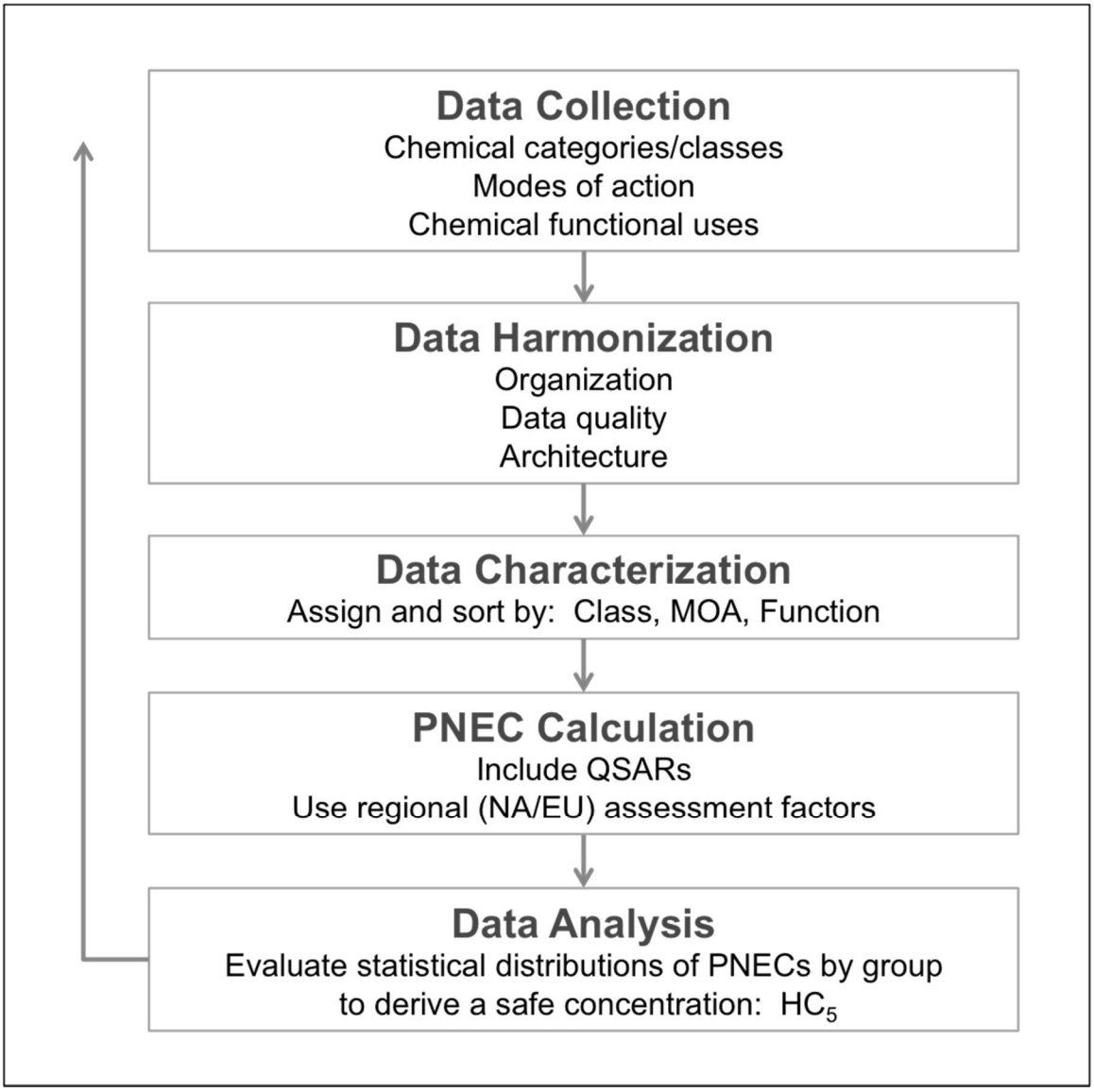

Figure 1. Process diagram showing the progression of ecoTTC construction and refinement.

Dataset characterization

The master dataset contained toxicity data from multiple sources, including peer-reviewed literature and REACH (Registration, Evaluation, Authorisation, and Restriction of Chemicals) submissions, spanning 40 years. Toxicity data were

incorporated for more than 1500 unique CAS numbers and over 2000 unique chemicals. 47 algal species, 2 invertebrate species, and 9 vertebrate species were represented. 111 classes based on ECOSAR designations were represented.

ecoTTC calculation

A subset of the master dataset was extracted at random for a preliminary calculation of a comprehensive ecoTTC (HC_5). This subset incorporated PNECs from 160 unique CAS numbers. Using North American application factor guidelines, the calculated HC_5s were 0.214 µg/L (most sensitive endpoint overall) (Figure 2) and 0.221 µg/L (most sensitive chronic endpoints only). Using EU application factor guidelines, the calculated HC_5s were 0.0561 µg/L (most sensitive endpoint overall) and 0.0544 µg/L (NA guidelines). HC_5 values were also calculated for a subgroup of 50 cationic chemicals. PNECs in this subgroup were calculated using North American application factor guidelines only. Using the most sensitive endpoint the HC_5 was 0.115 µg/L. Using the most sensitive chronic endpoints only the HC_5 was 0.113 µg/L.

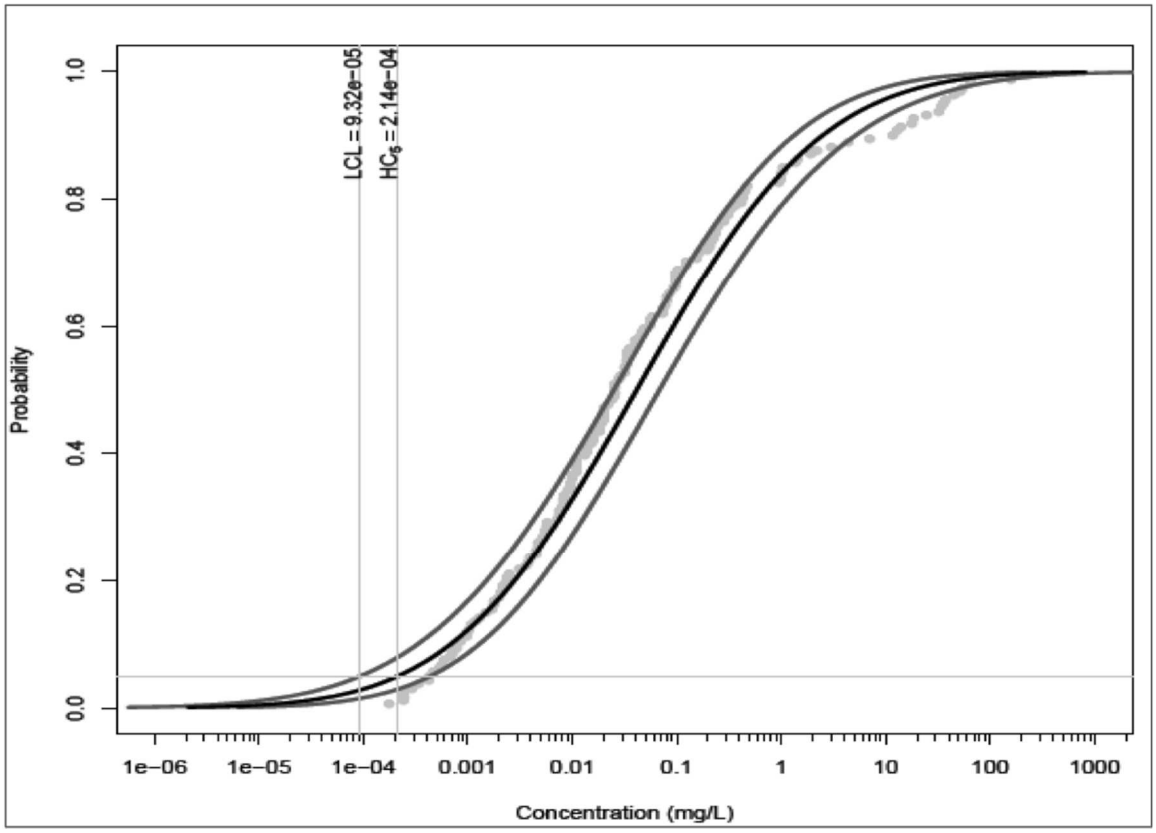

Figure 2. Distribution of a general subset of the master PNEC dataset. PNECs were calculated using North American application factor guidelines. HC_5 was calculated at 0.214 µg/L. *n* = 160.

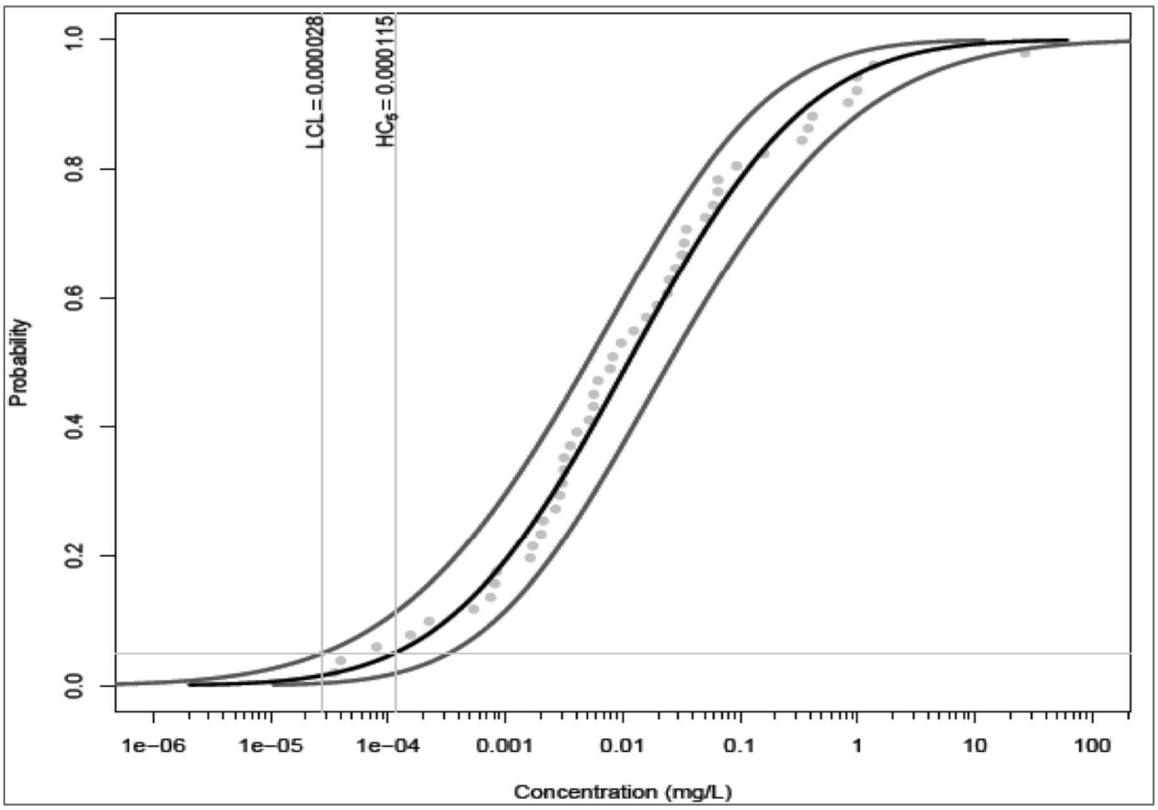

Figure 3. Distribution of a subset of cationically charged chemicals from the master PNEC dataset. PNECs were calculated using the overall most sensitive endpoints and North American application factor guidelines. HC_5 was calculated at 0.115 µg/L. *n* = 50.

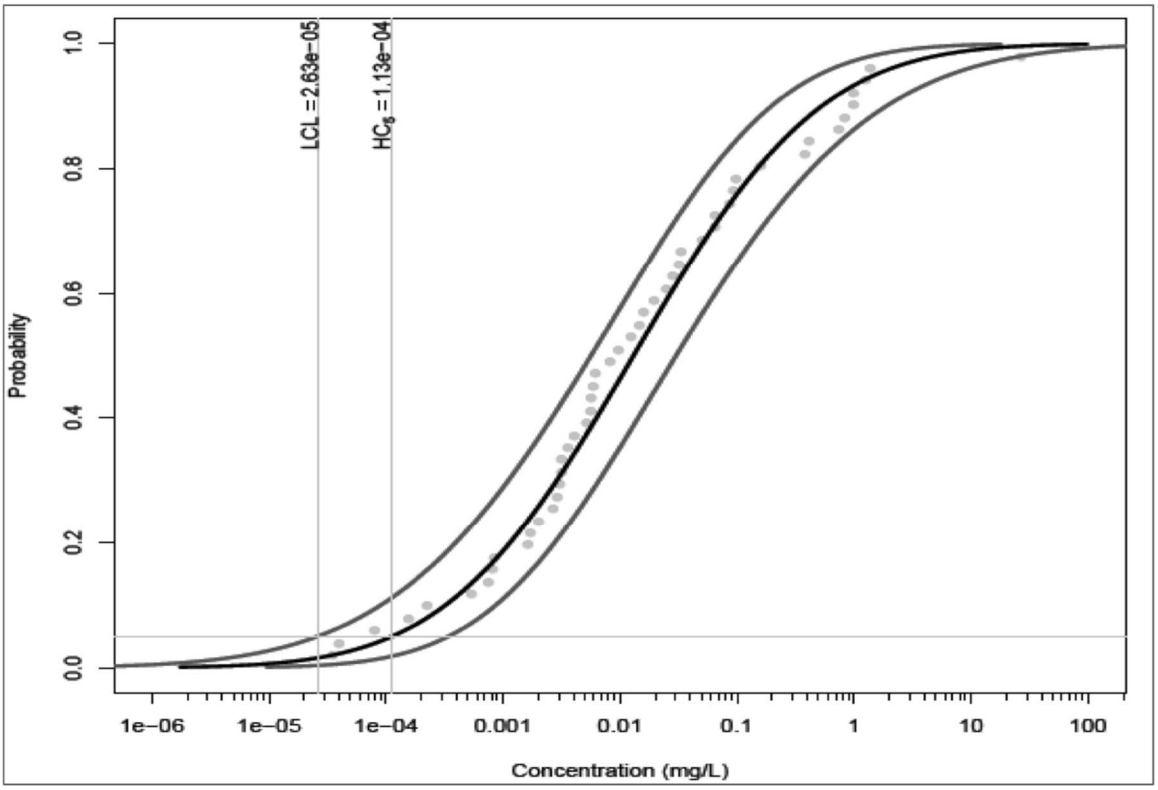

Figure 4. Distribution of a subset of cationically charged chemicals from the master PNEC dataset. PNECs were calculated using the most sensitive chronic endpoints only and North American application factor guidelines. HC_5 was calculated at 0.113 µg/L. *n* = 50.

DISCUSSION

'Threshold of toxic concern' is not novel; this concept has been in use for decades in human health and has been widely adapted to cover many applications[4, 8, 15]. With the further implementation of REACH registration requirements and the ever-increasing pace of new chemical technology, the TTC concept has become attractive to risk assessors and regulators searching for faster, cheaper, better ways to screen chemicals for toxicity. An ecotoxicological

TTC would be especially useful for the many chemicals produced at low tonnage or those for which little toxicology data exists.

Although the TTC has been applied to certain subcategories of chemical class or route of exposure, much is yet to be understood about the scope and parameters of its overall application. For instance, the choice of data to include in the initial data collection step and the relevant endpoints vary widely between studies [8-10]. Previous applications of similar ecoTTC concepts used a variety of statistical methods to construct threshold values. In this study, each subset of data for a unique CAS number and chemical was compiled and from that data a PNEC calculated. PNECs included application factors and ECOSAR modeling data in order to reflect an appropriate amount of inherent uncertainty in the final threshold value. A lognormal distribution of PNECs was used to determine the HC_5, which was chosen to represent a 'safe' concentration and the final ecoTTC.

Parameter choices within the PNEC calculation were expected to influence final TTC values. OECD regulatory guidelines only support the use of chronic data endpoints to construct PNECs, while American guidelines allow incorporation of acute data [12, 13]. Regional guidelines also differ widely in the use of application factors to manage uncertainty. In order to better understand these influences, PNECs were calculated from a representative subset of data using two different endpoint approaches and two sets of regional application factors. Using the most sensitive endpoints overall (inclusion of acute and chronic data)

and the North American (NA) application factor guidelines, the calculated HC_5 and therefore ecoTTC was 0.214 µg/L (Figure 3). OECD European (EU) application factor guidelines returned an ecoTTC of 0.0561 µg/L. When using only chronic endpoints and NA application factors in the ecoTTC calculation, the HC_5 was 0.221 µg/L. EU application factors and chronic endpoints returned an ecoTTC of 0.0544 µg/L. These values correspond to threshold values previously proposed for a broad chemical dataset [10] and for organic chemicals only [9].

Analysis of the distribution of PNEC values in this study showed that cationically charged chemicals were usually positioned in the low end of the distribution. Modeling a subset of 50 cationics produced an HC_5 of 0.115 µg/L (Figure 3). Similarly, using chronic endpoints only, the cationic subset returned an HC_5 of 0.113 µg/L (Figure 4). These preliminary ecoTTC values were roughly an order of magnitude higher than values proposed for cationics in Gutsell et al. (2015), confirming the need for further exploration of key chemical classes in ecoTTC development [10].

CONCLUSION

In this study, a metadataset of ecotoxicity data covering over 1500 chemicals was constructed using peer-reviewed literature and publicly available chemical registration data through REACH. PNECs were calculated for a subset of chemicals that included data-poor and difficult-to-test chemicals. Preliminary analyses indicated that the ecoTTC from a subset of cationically charged

chemicals was significantly higher than that of Gutsell et al. [10]. Future work includes complete characterization of the metadataset, along with PNEC calculations for each unique CAS/chemical. The ecoTTC values will be further analyzed to understand consequences of structure and mode of action on overall toxicity and the ability to extrapolate those values to unknown chemicals.

REFERENCES

[1] Kroes R, Kozianowski G. 2002. Threshold of toxicological concern (TTC) in food safety assessment. *Toxicology letters* 127:43-46.
[2] Munro IC, Ford RA, Kennepohl E, Sprenger J. 1996. Correlation of structural class with no-observed-effect levels: a proposal for establishing a threshold of concern. *Food and chemical toxicology* 34:829-867.
[3] Dolan DG, Naumann BD, Sargent EV, Maier A, Dourson M. 2005. Application of the threshold of toxicological concern concept to pharmaceutical manufacturing operations. *Regulatory toxicology and pharmacology* 43:1-9.
[4] Barlow S, Kozianowski G, Würtzen G, Schlatter J. 2001. Threshold of toxicological concern for chemical substances present in the diet. *Food and chemical toxicology* 39:893-905.
[5] Clements RG, Nabholz J, Zeeman M. 1996. *Estimating toxicity of industrial chemicals to aquatic organisms using structure activity relationships*. US EPA, Office of Pollution Prevention and Toxics.
[6] Carthew P, Clapp C, Gutsell S. 2009. Exposure based waiving: the application of the toxicological threshold of concern (ttC) to inhalation exposure for aerosol ingredients in consumer products. *Food and chemical toxicology* 47:1287-1295.
[7] Kroes R, Galli C, Munro I, Schilter B, Tran L-A, Walker R, Würtzen G. 2000. Threshold of toxicological concern for chemical substances present in the diet: a practical tool for assessing the need for toxicity testing. *Food and chemical toxicology* 38:255-312.
[8] Gross M, Daginnus K, Deviller G, de Wolf W, Dungey S, Galli C, Gourmelon A, Jacobs M, Matthiessen P, Micheletti C. 2010. Thresholds of toxicological concern for endocrine active substances in the aquatic environment. *Integrated environmental assessment and management* 6:2-11.
[9] de Wolf W, Siebel-Sauer A, Lecloux A, Koch V, Holt M, Feijtel T, Comber M, Boeije G. 2005. Mode of action and aquatic exposure thresholds of no concern. *Environmental toxicology and chemistry* 24:479-485.
[10] Gutsell S, Hodges G, Marshall S and Roberts J. (2015), Ecotoxicological thresholds—practical application to an industrial inventory. Environmental Toxicology and Chemistry, 34: 935–942.
[11] USEPA. 2014. Estimation Programs Interface Suite™ for Microsoft® Windows. 4.11 ed. United States Environmental Protection Agency, Washington, DC, USA.
[12] Nabholz J, Miller P, Zeeman M. 1993. Environmental risk assessment of new chemicals under the Toxic Substances Control Act TSCA Section Five. *ASTM SPECIAL TECHNICAL PUBLICATION* 1179:40-40.
[13] OECD. 2011. *Manual for the Assessment of Chemicals, Chapter 4: Initial Assessment of Data*. OECD Publishing.
[14] R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[15] Blackburn K, Stickney JA, Carlson-Lynch HL, McGinnis PM, Chappell L, Felter SP. 2005. Application of the threshold of toxicological concern approach to ingredients in personal and household care products. *Regulatory toxicology and pharmacology* 43:249-259.

PROJECT CONCLUSION

In this study, the Embryonic Developmental Rate Assay was designed to use changes to developmental timing in zebrafish embryogenesis as a simple and useful biomarker of toxicity. Developmental timing was marked by morphological staging and EED measurements and each measured delay consistently. The assay was tested using embryos treated with known toxicants copper and TDS, which showed differences in developmental rate that were not significant.

The Strategic Information-Filtering Tool was developed for use with large datasets typically used for risk assessment to allow probing for analytical purposes other than risk assessment. The tool used a stepwise technique to winnow down data while evaluating for reliability and relevance. Application of the tool in a case study of chronic toxicity data analysis and a comparison to the application of two similar tools showed the utility of SIFT for this purpose.

Use of the NOEC versus ECx summary statistic has engendered controversy; however, a comprehensive, literature-based evaluation of the NOEC:ECx relationship had not been accomplished. This study of chronic *Daphnia* test data showed that the EC_{10} was a strong analogue for the NOEC, and that key test parameters impacted the strength of that relationship. Refinements to test guidelines and future evaluation of other water quality parameters were recommended.

The Threshold of Toxic Concern (TTC) concept has been successfully applied in the human safety context, but has yet to fully transfer to the ecotoxicological context. In this study, construction and analysis of a potential eco-TTC framework was accomplished, resulting in a set of values consistent with other, similar approaches [12,13]. Evaluations of test parameters, chemical class, and application factors revealed that all influenced final ecoTTC values.

In the current regulatory environment, where animal welfare and chemical management are global priorities, new approaches to animal toxicity testing are vital. Equally important are the harmonization and optimization of existing test methods. This study contributes to these needs by introducing novel methodologies and evaluating current test guidelines from a foundation of 'replacement, refinement, and reduction'.

REFERENCES

[1] European Parliament, The Council of the European Union. 2006. Regulation (EC) no. 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the registration, evaluation, authorization and restriction of chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) no. 793/93 and Commission Regulation (EC) no. 1488/94, as well as Council Directive 76/769/EEC and commission Directives 91/155/EEC, 93/67/EEC, 93/105/CE and 2000/21/EC. *Official journal of the European Union* 396.

[2] OECD. 2012. *FISH TOXICITY TESTING FRAMEWORK Series on Testing and Assessment No. 171* OECD Publishing.

[3] Zeeman M. 1995. EPA's framework for ecological effects assessment. *Screening and testing chemicals in commerce September 1995* 23:32.

[4] OECD. 2004. *New Chemical Assessment Comparisons and Implications for Work Sharing, No. 48*. OECD Publishing.

[5] 1976. Toxic Substances Control Act (TSCA), Public Law 94-469, 90 United States.

[6] Russell WMS, Burch RL, Hume CW. 1959. The principles of humane experimental technique.

[7] OECD. 2013. *Test No. 236: Fish Embryo Acute Toxicity (FET) Test*. OECD Publishing.

[8] OECD. 1992. *Test No. 210: Fish, Early-Life Stage Toxicity Test*. OECD Publishing.

[9] OECD. 2012. *Test No. 229: Fish Short Term Reproduction Assay*. OECD Publishing.

[10] Parliament E. 2013. REPORT FROM THE COMMISSION TO THE COUNCIL AND THE EUROPEAN PARLIAMENT Seventh Report on the Statistics on the Number of Animals used for Experimental and other Scientific Purposes in the Member States of the European Union.

[11] Embry MR, Belanger SE, Braunbeck TA, Galay-Burgos M, Halder M, Hinton DE, Léonard MA, Lillicrap A, Norberg-King T, Whale G. 2010. The fish embryo toxicity test as an animal alternative method in hazard and risk assessment and scientific research. *Aquatic toxicology* 97:79-87

[12] Gutsell S, Hodges G, Marshall S and Roberts J. (2015), Ecotoxicological thresholds—practical application to an industrial inventory. Environmental Toxicology and Chemistry, 34: 935–942.

[13] de Wolf W, Siebel‐Sauer A, Lecloux A, Koch V, Holt M, Feijtel T, Comber M, Boeije G. 2005. Mode of action and aquatic exposure thresholds of no concern. *Environmental toxicology and chemistry* 24:479-485.