



The utility of behavioral studies for aquatic toxicology testing: A meta-analysis



Steven D. Melvin*, Scott P. Wilson

Central Queensland University, Centre for Environmental Management, Gladstone, Queensland 4680, Australia

HIGHLIGHTS

- There has been recent interest in behavioral responses for aquatic toxicity testing.
- We located behavioral, acute lethality, developmental and reproductive studies.
- Using meta-analysis we compared sensitivity, duration, effect size, and power.
- Behavioral studies represent fast, sensitive, and powerful tools for toxicology.
- Increased focus developing and optimizing behavioral tools would be valuable.

ARTICLE INFO

Article history:

Received 15 March 2013
Received in revised form 12 July 2013
Accepted 15 July 2013
Available online 16 August 2013

Keywords:

Aquatic toxicology
Behavioral analyses
Meta-analysis
Sensitivity
Experimental methodology

ABSTRACT

Behavioral responses have been applied for decades as tools for aquatic toxicity testing, but have received far less attention than studies assessing lethality, development or reproduction. With improved visual and non-visual assessment tools and increased knowledge of the importance of behavior for organism health and fitness, interest in behavioral analysis has increased in recent years. However, to our knowledge there has never been a quantitative assessment of the available techniques for organismal toxicity testing, so it is not clear whether behavioral studies represent valuable additions to environmental monitoring. We performed a meta-analysis comparing the relative sensitivities and average durations of behavioral studies to those assessing acute lethality, development and reproduction. Results demonstrate that the average duration of behavioral studies is consistently less than developmental or reproductive studies, and that behavioral endpoints are generally more sensitive than those assessing development or reproduction. We found effect sizes to be lower but power to be higher in behavioral and reproductive studies compared to studies assessing development, which likely relates to low sample sizes commonly used in developmental studies. Overall, we conclude that behavioral studies are comparatively fast and sensitive, and therefore warrant further attention as tools for assessing the toxicological effects of environmental contaminants. We suggest that research aimed at developing and optimizing techniques for behavioral analysis could prove extremely useful to the field of toxicology, but that future work must be directed at determining what specific behaviors are most sensitive to various classes of contaminants, and at understanding the relevance of changes to discrete behaviors for influencing organismal and population-level health and fitness.

© 2013 The Authors. Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Behavioral responses of aquatic organisms have been used for decades as methods for environmental monitoring (Cairns and Gruber, 1980; Kramer et al., 1989; Diamond et al., 1990; Gerhardt

et al., 1998; van der Schalie et al., 2001), but these types of studies have previously received much less attention than areas such as developmental or reproductive toxicology (Scott and Sloman, 2004). This is largely due to the absence of user-friendly tools facilitating image acquisition or other behavioral endpoints, but also relates to our limited understanding of the natural behaviors of many organisms or the relevance of behavioral responses for inferring higher-level effects such as survival, development and organismal fitness (Kane et al., 2005). However, in recent years considerable advancements have been made in the technological tools available for quantifying behavior (Lv et al., 2013), and our understanding of relationships between animal behaviors and

* Corresponding author. Tel.: +61 749 707 609; fax: +61 749 707 207.
E-mail address: s.d.melvin@cqu.edu.au (S.D. Melvin).

physiological and ecological consequences is improving (Little and Brewer, 2001; Amiard-Triquet, 2009; Sloman and McNeil, 2012). This has led to newfound interest into the use of behavior studies for aquatic toxicology testing, and therefore calls for an assessment of the utility of behavioral studies compared to other more established methodologies assessing the organismal level effects of environmental contaminants.

Toxicological studies have conventionally focused on assessing acutely lethal concentrations (e.g., median lethal concentration, LC50) and chronic sub-lethal effects on developmental or reproductive endpoints (Hood, 2005; Stadler, 2011), since these techniques provide results that can be directly related to organismal health and fitness. However, depending on the life history and developmental characteristics of different model species, assessing chronic endpoints can require a substantial amount of time and costs. With the wide range of environmental contaminants finding their way into aquatic environments there is a growing need for monitoring tools that are fast and sensitive to a wide range of compounds, but also indicative of potential effects on survival, growth and fitness. Behavioral analyses show promise for satisfying these requirements, and are often hailed for their rapidity (Gerhardt, 2007a; Maradona et al., 2012) and sensitivity (Miller et al., 1982; Little and Finger, 1990; Cleveland et al., 1991; Little et al., 1993) compared to traditional toxicological methods assessing developmental and reproductive effects. Although few would argue that behavioral responses offer comparatively fast and sensitive assessment of environmental perturbations (Gerhardt, 2007a; Amiard-Triquet, 2009), this has never been explicitly investigated on a broad range of contaminants or amongst different model organisms.

The use of meta-analyses to quantitatively synthesize ecological and toxicological data has increased in recent years, since these techniques allow broader questions to be investigated than conventional experimental hypothesis testing (Osenberg et al., 1999). The most common application of meta-analyses is to compare effect sizes within-studies in order to assess heterogeneity in the data amongst a set of conditions (e.g., is growth affected by exposure to a certain chemical?). A somewhat different approach that has been successfully applied to ecological data is the comparison of some common outcome amongst different study-types (Melvin and Houlahan, 2012). With this approach, it is not necessary for measures of effect size to be calculated for each individual study (i.e., comparison of within-study effects), because the focus is on how the study-type influences the outcome (i.e., comparisons amongst study-types). Consequently, this type of meta-analysis is well suited to investigating how different types of toxicological studies compare in their sensitivities and overall usefulness for assessing the effects of environmental contaminants.

The utility of behavioral analyses for assessing the adverse consequences of aquatic contaminants needs to be considered, in order to ascertain whether the further development and standardization of behavioral testing procedures would offer meaningful additions to the field of ecotoxicology. To this end, we performed a systematic review and meta-analysis of the literature to compare studies assessing behavioral responses to more established acute-lethality, developmental and reproductive procedures in order to answer the following questions:

- (1) Are the average timeframes of behavioral experiments comparable to those of acute-lethality, developmental and reproductive experiments?
- (2) Are behavioral experiments comparable in their sensitivity to acute-lethality, developmental and reproductive experiments?
- (3) Are response magnitudes (effect sizes) comparable amongst studies assessing behavioral, developmental, and reproductive endpoints?

- (4) Does the statistical power achieved with behavioral studies compare to that of studies assessing developmental and reproductive endpoints?

2. Materials and methods

We performed an extensive search of the literature for as many studies as we could identify describing behavioral responses of aquatic organisms to environmental contaminants. We focused only on studies describing aqueous exposures, because exposure route has been shown to influence behavioral responses (Gerhardt, 2007b). Our literature search was initially focused on behavioral studies to establish the list of contaminants and test organisms to be included in the review. We searched the Web of Knowledge™ and Science Direct™ referencing databases using combinations of the following search terms: 'Behavioral Early Warning System', 'BEWS', 'behavioral', 'behavior', 'analysis', 'swimming', 'movement', 'toxicity', 'exposure', and sorted the results by relevance to these search terms. For each combination of search terms we limited our search to the first 150 publications returned. However, in order to ensure a comprehensive dataset we also searched the references section of each article in an attempt to identify additional relevant behavioral studies. Our criteria for including a study was that it (1) described the waterborne exposure of an environmental contaminant to an aquatic organism, (2) clearly described the species and contaminants included in the study, (3) described the specific endpoints being assessed (4) reported the concentrations at which significant behavioral responses were observed, and (5) stated the duration of time required for the study.

Following our initial search, we performed a subsequent search of the literature to locate studies describing acute-lethality (LC50) estimates, and developmental and reproductive endpoints for the same combinations of contaminants and classes of organisms described in the behavioral studies. With this focus, we systematically searched the referencing databases with various search terms combining (i) the class of organism, (ii) the contaminant of interest, and (iii) the endpoint of interest/study type. For example, if a behavioral study described the effects of contaminant X on species Y, we searched the referencing databases for: 'X Y LC50', 'X Y acute lethal', 'X Y growth', 'X Y growth effect', 'X Y development', 'X Y developmental effect', 'X Y reproduction', and 'X Y reproductive effect'. As with the behavioral studies, we sorted the search results by relevance and limited our search to the first 150 publications returned. Our criteria for including a study was the same as described for the behavioral search, with the exception of acute-lethality studies where the only requirement was that an LC50 (24-, 48-, 72-, or 96-h) concentration was provided for the same combinations of contaminant and organism identified with the initial search of the behavioral literature. While we recognize that the timeframe for studies assessing acute-lethality could influence the resulting LC50 estimates, we chose not to control for this in our data analyses for two reasons – First, differences in LC50 estimates from 24- or 96-h tests would be relatively small in comparison to the differences in sensitivity between studies assessing acute-lethality and those investigating behavioral, developmental and reproductive endpoints. Second, if we assume that 24-h exposures yield the lowest LC50 estimates then including longer timeframes would make comparisons of acute-lethality studies to behavioral, developmental and reproductive studies more conservative rather than exaggerated, thus adding weight to our results.

For all papers that met the inclusion criteria following both the initial (i.e., behavioral) and subsequent (i.e., LC50, developmental and reproductive) literature searches we recorded the class of organism used (e.g., fish, crustacean, bivalve, etc.), the contaminant studied, the lowest concentration reported as eliciting a significant response (Minimum Response Concentration, MRC; mg L), and the

duration of the study (d). Once the complete data set was compiled, we sorted by contaminant and filtered by class of organism, then excluded toxicants where information did not exist for all four types of study (i.e., behavioral, LC50, developmental and reproductive) to facilitate comparison of sensitivities amongst study types. The only studies that met this criterion were those with fish and crustaceans as the model organisms. To compare the relative sensitivities of the various types of toxicological studies, we calculated the standardized sensitivity (z-score) of all MRCs independently for each contaminant. z-scores were calculated for fish and crustaceans separately using the formula,

$$z = \frac{x - \mu}{\sigma}$$

where x is the individual raw MRC data point, μ is the population mean of all MRCs for studies describing the same contaminant and class of organism, and σ is the standard deviation (SD) of all MRCs for studies describing the same contaminant and class of organism. To ensure that our analysis and interpretations were robust we also compared z-scores for all contaminants where data was only available for three study-types, provided at least one of these was a behavioral study.

Where the data was provided in a study, we recorded the number of replicates (N), mean of the response variable (e.g., average swim speed, developmental rates, reproductive output) and associated measure of error (i.e., standard deviation (SD) or standard error (SEM)) for the control group and the MRC treatment group. We used this data to calculate the pooled standard deviation,

$$SD_{pooled}^* = \sqrt{\frac{(n_c - 1)SD_c^2 + (n_t - 1)SD_t^2}{n_c + n_t - 2}}$$

where n_c and SD_c are the sample size and standard deviation of the control group, and n_t and SD_t are the sample size and standard deviation of the MRC treatment group, respectively. We then calculated the Hedges' g effect size between the control and MRC treatment groups of each of these studies according to the formula,

$$Hedges' g = \frac{M_c - M_t}{SD_{pooled}^*}$$

where M_c is the mean of the control group and M_t is the mean of the MRC treatment group. Finally, the pooled standard deviation and Hedges' g effect size were used to calculate the statistical power (1- β error probability) achieved by each study, using the freely available G*Power statistical software (Erdfeider et al., 1996). We analysed for differences in study duration, sensitivity, Hedges' g effect size, and statistical power between the various study types using randomizations (without replacement). This technique is well recognized as an appropriate and effective method for detecting differences in weighted-mean effect sizes amongst classes of studies (Adams et al., 1997; Gurevitch and Hedges, 1999; Melvin and Houlahan, 2012). Randomizations were done in MS EXCEL (Microsoft Inc.) by first calculating the sum of the square deviations of each group (study type) from the population mean and then randomly reassigning data points to the groups and re-computing the deviations. We used the Macro function and performed 10000 permutations of the data to determine the probability of randomly observing deviations from the mean greater than or equal to those related to the grouping variable (i.e., study type), with $\alpha = 0.05$ as the significance threshold.

3. Results

Our initial search identified 90 studies describing 163 individual behavioral responses of aquatic organisms to various contaminants. These studies included data for 52 species and 68 different

contaminants – the majority of species (86%) were fishes (45%), crustaceans (28%) and bivalves (13%) and the contaminants (66%) mainly included pesticides (33%), metals (25%), and pharmaceuticals (8%). Our subsequent search for experiments assessing the same classes of organisms and contaminants identified 88 studies describing 96 developmental responses, 73 studies describing 81 reproductive responses, and 155 studies reporting 324 LC50 concentrations. After sorting the full data set by contaminant and filtering by organism class, fish and crustaceans were the only groups where data was available describing the same contaminant(s) for all four types of study. This left us with 106 data points from 60 studies describing 11 different contaminants for crustaceans, and 220 data points from 133 studies describing 14 contaminants for fish (Electronic Supplementary Materials 1 and 2).

The timeframe (days \pm SD) required for behavioral studies (1.58 \pm 5.25) and those assessing acute-lethality (3.46 \pm 1.02) averaged significantly lower ($p < 0.001$) than studies assessing developmental (43.54 \pm 68.03) or reproductive endpoints (44.34 \pm 69.45). Duration was much more variable for studies assessing developmental and reproductive endpoints than those assessing behavior or acute-lethality (Fig. 1). Study duration can represent a major source of confounding in meta-analyses (Osenberg et al., 1999) and data often needs to be corrected to account for this (e.g., Melvin and Houlahan, 2012). However, this was not a concern for our analysis since the goal was to compare sensitivities of various study types with very different outcomes, and differences in duration are an inevitable consequence of different study types.

Acute-lethality estimates (LC50s) were the least sensitive parameter overall ($p < 0.001$) and this was also true when considering studies with either crustaceans ($p = 0.026$) or fish ($p < 0.001$) as the model organism independently (Fig. 2). Relative to studies investigating acute-lethality, behavioral studies were comparatively more sensitive overall than those assessing developmental and reproductive endpoints. For studies with crustaceans as the model organism, there was no statistical difference in the comparative sensitivity of behavioral, developmental or reproductive studies. Conversely, behavioral responses had greater sensitivity than those investigating developmental ($p = 0.054$) or reproductive ($p = 0.017$) endpoints in studies with fish as the model organism (Fig. 2). Although the sensitivity of behavioral studies was comparatively greater on average than studies assessing both developmental and reproductive endpoints, this was not true for every contaminant. Behavioral studies had the lowest sensitivities for some contaminants with both crustaceans and fishes (Tables 1–3). For example, behavioral studies with crustaceans were less sensitive than developmental or reproductive studies for cetyltrimethylammonium bromide, lindane, and zinc (Table 1). Similarly,

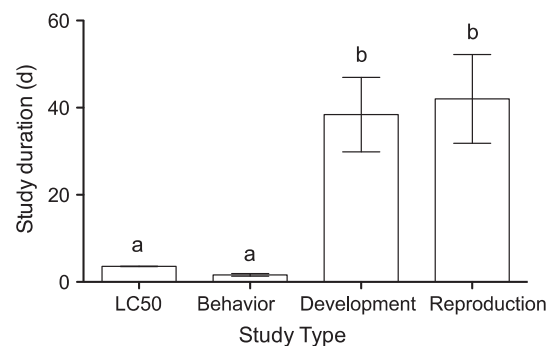


Fig. 1. Mean duration of studies assessing acute-lethality, behavioral, developmental and reproductive endpoints. Data is for the comprehensive list of publications compiled through searches of referencing databases. Bars represent \pm 1 SE. Different letters indicate significant differences ($\alpha < 0.05$) based on 10000 random permutations of the data.

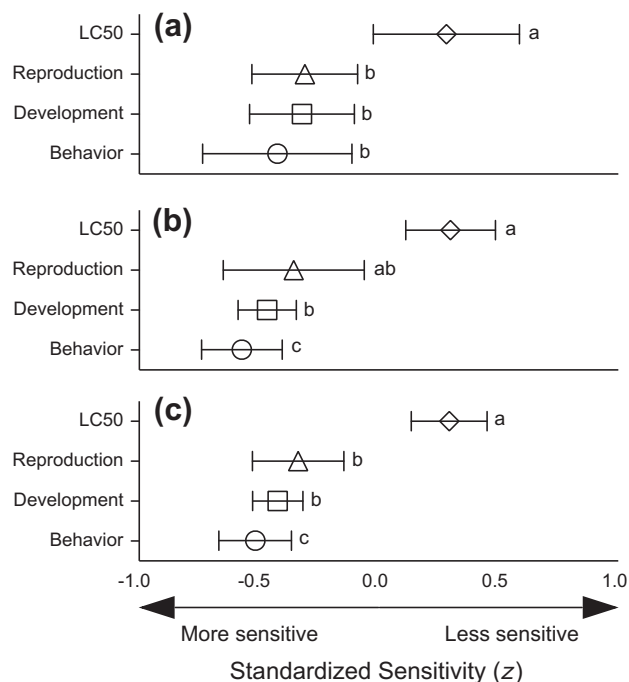


Fig. 2. Mean sensitivity scores (z-scores) of studies with (a) crustaceans and (b) fish as the model organisms, and (c) overall z-scores, for studies assessing acute lethality, behavioral, developmental and reproductive endpoints. z-scores were calculated independently for each contaminant where information existed for all study types. Bars represent 95% confidence interval. Different letters indicate significant differences ($\alpha < 0.05$) based on 10000 random permutations of the data.

behavioral studies with fish were less sensitive than developmental or reproductive studies for cyanide and mercury (Table 2). Differences in the comparative sensitivities of the various study types were more evident when considering contaminants where data was only available for three study-types (Table 3).

The average effect size of behavioral studies was approximately 4× lower than that of developmental studies and 2× lower than studies assessing reproductive endpoints ($p = 0.002$; Fig. 3). Although effect sizes were on average the greatest in developmental studies, the statistical power of developmental studies averaged lower than both behavioral and reproductive studies (Fig. 3). However, this difference was not statistically significant ($p = 0.062$). Since sample size is an important contributor to both effect size and power, it is important to note that the mean numbers of replicates reported for the behavioral, developmental, and reproductive studies included in these calculations were 12.5, 5.5, and 9.8, respectively.

Table 1
Average standardized sensitivity scores (z-scores) of studies describing crustaceans, for contaminants with information available for all study types. References refer to the list of publications provided in [Electronic Supplementary Material 1](#).

| Contaminant | LC50 | Behavior | Development | Reproduction | References |
|--------------------------------|--------|----------|-------------|--------------|--|
| Atrazine | 0.153 | -0.684 | 0.199 | -0.658 | 24,86,138,178,183,190,192 |
| Cetyltrimethylammonium bromide | 0.094 | 0.530 | -0.671 | -0.671 | 39,47,57,65,118 |
| Chlorpyrifos | 0.495 | -0.394 | -0.396 | -0.397 | 6,33,82,109,134 |
| Copper | 0.243 | -0.332 | -0.335 | -0.333 | 8,10,21,22,48,64,73,94,150,155,167,183 |
| Cypermethrin | 0.167 | -0.333 | -0.333 | -0.333 | 3,7,34,40,66,116,182 |
| Deltamethrin | 0.623 | -0.740 | -0.687 | -0.442 | 14,16,66,116 |
| Ibuprofen | 1.072 | -0.798 | -0.179 | -0.092 | 31,47,57,62,111,137,151 |
| Lindane | 0.388 | -0.265 | -0.332 | -0.355 | 4,11,91,97,98,145,149,168,183 |
| Mercury | 0.266 | -0.560 | -0.605 | -0.433 | 21,25,100,145,167,184 |
| Paraoxon-methyl | -0.552 | -1.131 | 0.842 | 0.842 | 51,85,107 |
| Zinc | 0.591 | -0.631 | -0.839 | -0.852 | 89,93,155,165,192 |
| Total | 3.540 | -5.338 | -3.336 | -3.724 | |

4. Discussion

Should the analysis of behavioral responses play a larger role in aquatic toxicity testing? The results of our meta-analysis suggest that behavioral responses might indeed represent favorable endpoints for assessing organismal and ecological effects of environmental contaminants. Comparisons of study duration, sensitivity, effect size and statistical power all indicate that behavioral responses should be more widely utilized and suggest that further research focused on optimizing and standardizing protocols that assess behavior would be beneficial to the field of aquatic toxicology. However, we stress that for behavioral studies to offer the most meaningful assessment of toxicological risk, research must be focused towards understanding how changes to discrete animal behaviors relate to broader ecological concerns such as survival, health and fitness. Furthermore, although behavioral studies reported the greatest sensitivity overall when compared to the other study types, this was not true for every contaminant included in our analysis. This also requires further investigation, in order to understand what specific behaviors are the most appropriate for assessing different classes of contaminants.

Our results demonstrate the extreme rapidity of studies assessing behavioral responses of aquatic organisms to a variety of known environmental contaminants. In the face of an ever-growing number of compounds with potential toxicological properties (Ritter et al., 2002; Fleeger et al., 2003; Lapworth et al., 2012), toxicity-testing procedures that offer consistently fast assessment of a wide range of contaminants are particularly desirable. Cell cultures and *in vitro* techniques are extremely useful for understanding mechanisms of toxicity and have been suggested for addressing this need (Shukla et al., 2010), but behavioral responses may provide information that is better connected to integrated whole-organism responses and broader ecological concerns. Therefore, the development of non-invasive strategies examining organismal responses to contaminants have the potential to be valuable components of toxicological testing regimes, because they may help assess true ecological consequences of exposure. There is growing evidence demonstrating the implications of organism behavior for influencing health and fitness through effects on survival, development and reproductive processes (Martinović et al., 2007; McGlynn, 2012; Denoël et al., 2013), but our results suggest that this needs to be investigated with a wider range of test organisms and exposure conditions.

Several behavioral parameters were affected in the studies included in our analysis, including for example, swim speed, distance moved, activity levels, spatial distribution patterns, feeding rates, and courting events. There are some very good examples of how effects on such behavioral parameters can influence organismal health and fitness. In the face of predation, evasive behaviors are

Table 2

Average standardized sensitivity scores (z-scores) of studies describing fish, for contaminants with information available for all study types. References refer to the list of publications provided in Electronic Supplementary Material 1.

| Contaminant | LC50 | Behavior | Development | Reproduction | References |
|-------------------------------|--------|----------|-------------|--------------|--|
| Aluminum | 0.571 | −1.286 | −1.284 | −1.282 | 23,104,119,139,166 |
| Arsenic | 0.275 | −1.307 | 0.259 | −0.026 | 9,42,127,140,173,187 |
| Atrazine | 0.786 | −0.687 | −0.683 | −0.652 | 12,32,35,87,88,96,113,147,178,185 |
| Cadmium | 0.692 | −0.576 | −0.519 | −0.523 | 15,50,52,53,54,71,95,103,162,170,171,189 |
| Carbaryl | 0.020 | −0.622 | 0.516 | −0.385 | 43,49,58,78,99,123,159,163,176,179,181 |
| Carbofuran | 0.272 | −0.341 | −0.385 | −0.222 | 26,45,49,55,59,90,122,125,129 |
| Chlorpyrifos | 0.248 | −0.491 | −0.258 | −0.744 | 17,27,32,46,75,76,80,92,120,129,132,133,153,177,181 |
| Cyanide | 0.691 | −0.227 | −0.343 | −0.428 | 5,36,38,56,60,70,131,135,146,148,152,158,164 |
| Diazinon | −0.044 | −0.793 | −0.394 | 3.120 | 18,19,28,29,30,32,37,49,67,69,72,105,115,124,154,169,172,174 |
| Ethionylestradiol | 1.196 | −0.399 | −0.398 | −0.399 | 1,2,44,68,74,83,84,106,108,112,114,117,128,156 |
| Mercury | 0.363 | −0.064 | −0.650 | −0.667 | 5,13,20,79,95,157,169 |
| Microcystin-LR | 0.277 | −0.684 | −0.671 | −0.684 | 63,101,121,126,141,175,180,191 |
| Pentachlorophenol | 0.465 | −1.177 | −0.974 | −1.064 | 71,81,102,130,142,161,188 |
| Perfluorooctanesulphonic acid | 1.042 | −0.363 | −0.416 | −0.340 | 41,61,77,110,136,143,144,160,186 |
| Total | 6.854 | −9.017 | −6.200 | −4.296 | |

Table 3

Average standardized sensitivity scores (z-scores) of studies describing bivalves, crustaceans, and fish, for contaminants with information available for three study types (provided information was available for at least one behavioral study). References refer to the list of publications provided in Electronic Supplementary Material 1.

| Class | Contaminant | LC50 | Behavior | Development | Reproduction | References |
|-------------|--------------|--------|----------|-------------|--------------|--|
| Bivalves | Chlorine | −0.500 | −0.500 | 1.499 | | 287,276,269,301 |
| | Copper | 2.256 | −0.269 | −0.640 | | 287,194,268,193,298,71,242,302 |
| Crustaceans | Fenitrothion | | 1.155 | −0.577 | −0.577 | 280,220 |
| | Fluoxetine | 0.984 | −0.764 | | −0.330 | 47,57,203,250,299,303 |
| | Tributyltin | | −0.566 | 1.155 | −0.588 | 145,235,282 |
| Fish | Cypermethrin | 0.119 | −0.370 | −0.349 | | 28,34,40,237,243,304,305,306,307,308,309,310 |
| | Deltamethrin | 0.194 | −0.351 | | −0.811 | 262,292,311,312,313,314,315 |
| | Endosulfan | 0.111 | 0.757 | | −0.987 | 255,228,316,317,318,319,320,321 |
| | Fenitrothion | −0.257 | 0.473 | −0.689 | | 199,263,291 |
| | Fluoxetine | 0.894 | −0.609 | | −0.589 | 208,260,299,300,322 |
| | Phenanthrene | 0.254 | −0.623 | 0.737 | | 214,251,285,323,324 |
| | RDX | 0.477 | 0.864 | −0.909 | | 210,246,281,325 |
| | Verapamil | 1.154 | −0.617 | −0.537 | | 231,232 |

a primary survival strategy for many organisms (Vamosi and Schluter, 2002; Bradley et al., 2013), and compounds that influence the evasive ability of an organism may therefore directly affect its ability to avoid predation. Feeding behaviors relate closely to growth and development and if disrupted by contaminant exposure (Pestana et al., 2007; Alonso et al., 2009) this could have consequences not only for development but also for sexual maturation and other energy intensive processes. Spawning and courtship behaviors are extremely important for reproductive success in many organisms, and these behaviors may be disrupted by exposure to various contaminants (Weber, 1993; Hammerschmidt et al., 2002; Oshima et al., 2003; Martinović et al., 2007). However, despite the existing evidence, our understanding of correlations between behavior and fitness is still quite limited (Gerhardt, 2007a). Research investigating these relationships would not only increase the significance of behavioral analysis for assessing toxicological effects, but would benefit our fundamental understanding of behavioral ecology (Slooman and McNeil, 2012).

Behavioral, developmental, and reproductive studies all yielded significant responses at concentrations well below reported LC50 values. This is perhaps not surprising, but it clearly demonstrates the value of all three types of study for assessing eco-toxicological effects of contaminant exposure at realistic exposure concentrations, since most contaminants only ever occur in natural aquatic environments at concentrations well below those causing acute lethality (Kolpin et al., 2002; De Lange et al., 2006; Houtman, 2010). However, the average sensitivity of behavioral studies was comparatively greater than studies assessing developmental or reproductive endpoints for both crustaceans and fish (albeit only

significantly so in fish). Regardless of statistical significance, we take this higher sensitivity and the relatively limited published literature as evidence that behavioral studies are not being utilized to their full potential for environmental monitoring, and should be considered as options for eco-toxicity testing. This is especially true when one considers the extremely short timeframes required for behavioral analyses. On the other hand, contrary to our overall results, some studies describe behavioral responses occurring at greater concentrations than those eliciting growth effects (Bryan et al., 1995; Passino-Reader et al., 1995). Differences in the comparative sensitivities of various endpoints are likely related to differences in experimental conditions, including the species and contaminant(s) being studied. Indeed, despite our overall results indicating greater sensitivity of behavioral responses, developmental and reproductive endpoints were more sensitive for some contaminants (Tables 1 and 2). Another explanation for this is that the mode of action of some contaminants may be very closely related to development or reproduction, but poorly related to factors influencing behavior. Therefore, a multi-endpoint approach may be the most appropriate for comprehensive aquatic toxicity testing, such as the assessment of unknown contaminants (e.g., complex effluents and wastewaters).

Our results indicate that the average effect size of behavioral studies is significantly lower than that obtained with studies assessing developmental or reproductive endpoints. Developmental studies in particular generally reported very large differences between control and MRC treatment groups. Interestingly, large effect sizes are somewhat conflicting with the low power achieved by developmental studies, but this can be explained by the smaller

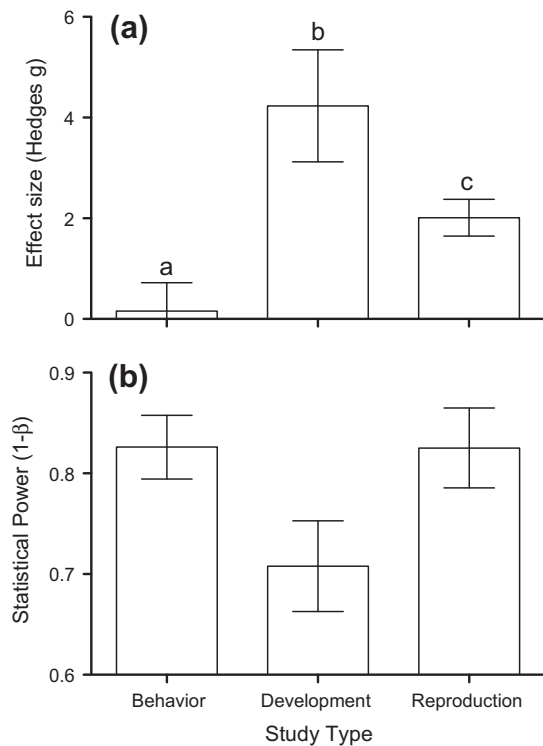


Fig. 3. Mean (a) Hedges' *g* effect size and (b) statistical power ($1-\beta$) of studies compiled through searches the referencing databases, where control and MRC treatment group samples sizes, means, and associated measures of variance were either provided or could be extracted. Bars represent $\pm 1SE$. Different letters indicate significant differences ($\alpha < 0.05$) based on 10000 random permutations of the data.

sample sizes generally utilized for developmental compared to behavioral or reproductive studies. Thus, differences in estimates of effect size and statistical power may be more reflective of trends in experimental set-up amongst the different study types, and these trends simply a reflection of the practicality of including large sample sizes in studies that are carried out over longer (i.e., developmental and reproductive studies) compared to shorter (i.e., behavioral studies) timeframes. The broad recommendation here, however, is that all studies use sufficient replication to ensure that statistical power is maximized (Steidl et al., 1997; Melvin et al., 2009). For studies assessing behavioral responses, achieving sufficiently large sample sizes is generally more feasible because the time commitment for these studies is often much shorter than that required for developmental or reproductive studies.

5. Conclusions

In conclusion, we have demonstrated and discussed the value of behavioral analyses as tools for assessing the toxicological consequences of environmental contaminants on aquatic organisms. Our study represents the first of its kind to quantitatively compare the duration, sensitivity, and statistical power of acute-lethality, behavioral, developmental, and reproductive studies through a comprehensive meta-analysis of the literature. We emphasize the fact that each of these study types provides meaningful data regarding the eco-toxicological effects of contaminant exposure on organismal and population parameters. However, considering the relatively high sensitivity and power achievable with behavioral studies in conjunction with the comparatively short timeframes they require, we advocate research aimed at optimizing and expanding on current behavioral techniques and the develop-

ment and standardization of innovative new procedures for behavioral analysis. Finally, in order for behavioral analysis to provide truly meaningful results relevant to ecological consequences of exposure, it is essential that research be directed at understanding linkages between specific behaviors and higher-level effects on survival, health and fitness.

Acknowledgements

We extend our gratitude to Ms. Chantal Lanctôt for reviewing earlier versions of the manuscript and providing constructive feedback. We acknowledge the Queensland Government Smart Futures Research-Industry Partnership Program for funding.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.chemosphere.2013.07.036>.

References

- Adams, D.C., Gurevitch, J., Rosenberg, M.S., 1997. Resampling tests for meta-analysis of ecological data. *Ecology* 78, 1277–1283.
- Alonso, Á., De Lange, H.J., Peeters, E.T.H.M., 2009. Development of a feeding bioassay using the freshwater amphipod *Gammarus pulex* and the multispecies freshwater biomonitor. *Chemosphere* 75 (3), 341–346.
- Amiard-Triquet, C., 2009. Behavioral disturbances: the missing link between sub-organismal and supra-organismal responses to stress? prospects bases on aquatic research. *Hum. Ecol. Risk Assess.* 15 (1), 87–110.
- Bradley, C.J., Strickler, J.R., Buskey, E.J., Lenz, P.H., 2013. Swimming and escape behaviour in two species of calanoid copepods from nauplius to adult. *J. Plankton Res.* 35 (1), 49–65.
- Bryan, M.D., Atchison, G.J., Sandheinrich, M.B., 1995. Effects of cadmium on the foraging behavior and growth of juvenile bluegill, *Lepomis macrochirus*. *Can. J. Fish. Aquat. Sci.* 52, 1630–1638.
- Cairns, J., Gruber, D., 1980. A comparison of methods and instrumentation of biological early warning systems. *Water. Resour. Bull.* 16 (2), 261–266.
- Cleveland, L., Little, E.E., Ingersoll, C.G., Wiedmeyer, R.H., Hunn, J.B., 1991. Sensitivity of brook trout to low pH, low calcium and elevated aluminum concentrations during laboratory pulse exposures. *Aquat. Toxicol.* 19, 303–317.
- De Lange, H.J., Noordoven, W., Murk, A.J., Lürling, M., Peeters, E.T.H.M., 2006. Behavioural responses of *Gammarus pulex* (Crustacea, Amphipoda) to low concentrations of pharmaceuticals. *Aquat. Toxicol.* 78, 209–216.
- Denoël, M., Libon, S., Kestemont, P., Brasseur, C., Focant, J.-F., De Pauw, E., 2013. Effects of a sublethal pesticide exposure on locomotor behavior: a video-tracking analysis in larval amphibians. *Chemosphere* 90, 945–951.
- Diamond, J.M., Parson, M.J., Gruber, D., 1990. Rapid detection of sublethal toxicity using fish ventilator behavior. *Environ. Toxicol. Chem.* 9, 3–11.
- Erdfelder, E., Faul, F., Buchner, A., 1996. A general power analysis program. *Behav. Res. Meth. Ins. Comput.* 28, 1–11.
- Fleeger, J.W., Carman, K.R., Nisbet, R.M., 2003. Indirect effects of contaminants in aquatic ecosystems. *Sci. Total. Environ.* 317, 207–233.
- Gerhardt, A., 2007a. Aquatic behavioral ecotoxicology—prospects and limitations. *Hum. Ecol. Risk Assess.* 13, 481–491.
- Gerhardt, A., 2007b. Importance of exposure route for behavioural responses in *Lumbriculus variegatus* Müller (Oligochaeta: Lumbriculida) in short-term exposures to Pb. *Environ. Sci. Pollut. Res.* 14 (6), 430–434.
- Gerhardt, A., Carlsson, A., Ressemann, C., Stick, K.P., 1998. A new online biomonitoring system for *Gammarus pulex* (L.) (Crustacea): in situ test below a copper effluent in South Sweden. *Environ. Sci. Technol.* 32, 150–156.
- Gurevitch, J., Hedges, L.V., 1999. Statistical issues in ecological meta-analyses. *Ecology* 80 (4), 1142–1149.
- Hammerschmidt, C.R., Sandheinrich, M.B., Wiener, J.G., Rada, R.G., 2002. Effects of dietary methylmercury on reproduction of fathead minnows. *Environ. Sci. Technol.* 36, 877–883.
- Hood, R.D., 2005. Developmental and reproductive toxicology: A practical approach, second ed. Taylor and Francis, Boca Raton, FL.
- Houtman, C.J., 2010. Emerging contaminants in surface waters and their relevance for the production of drinking water in Europe. *J. Integr. Environ. Sci.* 7 (4), 271–295.
- Kane, A.S., Salierno, J.D., Brewer, S.K., 2005. Fish models in behavioural toxicology: automated techniques, updates and perspectives. In: Ostrander, G.K. (Ed.), *Methods in Aquatic Toxicology*, vol. 2. Lewis Publishers, Boca Raton, FL, pp. 559–590.
- Kolpin, D.W., Furlong, E.T., Meyer, M.T., Thurman, E.M., Zaugg, S.D., Barber, L.B., Buxton, H.T., 2002. Pharmaceuticals, hormones, and other organic wastewater contaminants in US streams, 1999–2000: a national reconnaissance. *Environ. Sci. Technol.* 36, 1202–1211.

- Kramer, K.J.M., Jenner, H.A., Zwart, D.D., 1989. The valve movement response of mussels: a tool in biological monitoring. *Hydrobiologica* 188 (189), 433–443.
- Lapworth, D.J., Baran, N., Stuart, M.E., Ward, R.S., 2012. Emerging organic contaminants in groundwater: a review of sources, fate and occurrence. *Environ. Pollut.* 163, 287–303.
- Little, E.E., Brewer, S.K., 2001. Neurobehavioral toxicity in fish. In: Schlenk, D., Benson, W.H. (Eds.), *Target Organ Toxicity in Marine and Freshwater Teleosts New Perspectives: Toxicity and the Environment*, vol. 2. Taylor and Francis, London and New York, pp. 139–174.
- Little, E.E., Finger, S.E., 1990. Swimming behaviour as an indicator of sublethal toxicity in fish. *Environ. Toxicol. Chem.* 9, 13–19.
- Little, E.E., Fairchild, J.F., DeLonay, A.J., 1993. Behavioral methods for assessing the impacts of contaminants on early life stage fishes. In: Fuiman, L. (Ed.), *American Fisheries Society Symposium 14 Water Quality and the Early Life Stages of Fishes*. American Fisheries Society, Bethesda.
- Lv, Z., Tek, A., Da Silva, F., Empereur-mot, C., Chavent, M., Baaden, Marc., 2013. Game on, Science – How video game technology may help biologists tackle visualization challenges. *PLoS ONE* 8 (3), e57990. <http://dx.doi.org/10.1371/journal.pone.0057990>.
- Maradona, A., Marshall, G., Mehrvar, M., Pushchak, R., Laursen, A.E., McCarthy, L.H., Bostan, V., Gilbride, K.A., 2012. Utilization of multiple organisms in a proposed early-warning biomonitoring system for real-time detection of contaminants: preliminary results and modeling. *J. Hazard. Mater.* 219–220, 95–102.
- Martinović, D., Hogarth, W.T., Jones, R.E., Sorensen, P.W., 2007. Environmental estrogens suppress hormones, behaviour, and reproductive fitness in male fathead minnows. *Environ. Toxicol. Chem.* 26 (2), 271–278.
- McGlynn, T.P., 2012. How does social behaviour evolve? *Nature Educ. Knowl.* 3 (10), 69.
- Melvin, S.D., Houlahan, J.E., 2012. Tadpole mortality varies across experimental venues: do laboratory populations predict responses in nature? *Oecologia* 169, 861–868.
- Melvin, S.D., Munkittrick, K.R., Bosker, T., MacLachy, D.L., 2009. Detectable effect size and bioassay power of mummichog (*Fundulus heteroclitus*) and fathead minnow (*Pimephales promelas*) adult reproductive tests. *Environ. Toxicol. Chem.* 28 (11), 2416–2425.
- Miller, D.C., Lang, W.H., Graeves, J.O.B., Wilson, R.S., 1982. Investigations in aquatic behavioural toxicology using a computerized video quantification system. In: Pearson, J.G., Foster, R.B., Bishop, W.E. (Eds.), *Aquatic Toxicology and Hazard Assessment: Fifth Conference, STP 766*. American Society for Testing and Materials, Philadelphia, PA.
- Osenberg, C.W., Sarnelle, O., Cooper, S.D., Holt, R.D., 1999. Resolving ecological questions through meta-analysis: goals, metrics, and models. *Ecology* 80 (4), 1105–1117.
- Oshima, Y., Kang, I.J., Kobayashi, M., Nakayama, K., Imada, N., Honjo, T., 2003. Suppression of sexual behaviour in male Japanese medaka (*Oryzias latipes*) exposed to 17 β -estradiol. *Chemosphere* 50, 429–436.
- Passino-Reader, D.R., Berlin, W.H., Hickey, J.P., 1995. Chronic bioassays of rainbow trout fry with compounds representative of contaminants in Great Lakes fish. *J. Great Lakes Res.* 21, 373–383.
- Pestana, J.L., Ré, A., Nogueira, A.J., Soares, A.M., 2007. Effects of cadmium and zinc on the feeding behavior of two freshwater crustaceans: *Atyaephyra desmarestii* (Decapoda) and *Echinogammarus meridionalis* (Amphipoda). *Chemosphere* 68 (8), 1556–1562.
- Ritter, L., Solomon, K., Sibley, P., Hall, K., Keen, P., Mattu, G., Linton, B., 2002. Sources, pathways, and relative risks of contaminants in surface water and groundwater: a perspective prepared for the Walkerton inquiry. *J. Toxicol. Environ. Health A* 65 (1), 1–142.
- Scott, G.R., Sloman, K.A., 2004. The effects of environmental pollutants on complex fish behaviour: integrating behavioural and physiological indicators of toxicity. *Aquat. Toxicol.* 68, 369–392.
- Shukla, S.J., Huang, R., Austin, C.P., Xia, M., 2010. The future of toxicity testing: a focus on in vitro methods using a quantitative high-throughput screening platform. *Drug Discov. Today* 15 (23–24), 997–1007.
- Sloman, K.A., McNeil, P.L., 2012. Using physiology and behaviour to understand the responses of fish early life stages to toxicants. *J. Fish Biol.* 81, 2175–2198.
- Stadler, J., 2011. Developmental and reproductive toxicology. In: Lodolo, A., Stadler, J. (Eds.), *Pharmaceutical Toxicology in Practice: A Guide for Non-Clinical Development*. John Wiley & Sons, Inc., Hoboken, NJ, USA. doi:10.1002/9780470909911.ch9.
- Steidl, R.J., Hayes, J.P., Schaubert, E., 1997. Statistical power analysis in wildlife research. *J. Wildlife Manage.* 61 (2), 270–279.
- Vamosi, S.M., Schluter, D., 2002. Impacts of trout predation on fitness of sympatric sticklebacks and their hybrids. *Proc. R. Soc. London, B* 269, 923–930.
- Van der Schalie, W.H., Shedd, T.R., Knechtges, P.L., Widder, M.W., 2001. Using higher organisms in biological early warning systems for real-time toxicity detection. *Biosensors Bioelectron.* 16, 457–465.
- Weber, D.N., 1993. Exposure to sublethal levels of waterborne lead alters reproductive behaviour patterns in fathead minnow (*Pimephales promelas*). *Neurotoxicology* 14, 347–358.