UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL

Effects of Pharmaceuticals on the Reproduction of Aquatic Organisms: a Meta-Analysis

David Gil Valente Lopes

Mestrado em Bioestatística

Trabalho de Projeto orientado por:
Doutora Vanessa Fonseca
Professora Doutora Marília Antunes

2019

It's the job that's never started as takes longest to finish.

- J. R. R. Tolkien

Acknowledgements

First of all, I would like to thank Bernardo Duarte, without whom I would never have had the opportunity to do this work. Thank you for helping me give the first steps in this important stage of academic life.

I am very grateful to my thesis mentor, Vanessa Fonseca. Thank you for every hour, every message answered, every additional article sent for me to read and learn and thank you for guiding me through every step and your motivational support. Above all, thank you for providing me with a theme that combines both my graduation field (Biology) and masters' (Biostatistics).

Next, I would like to thank professor Marília Antunes for her valuable advises even when on a tight schedule. Every minute was helpful.

I would also like to extend my thanks to Mónica, who was very patient to listen to every little problem I had and to help me whenever she could.

Last but not least, I want to thank my parents, for being with me everyday, every step of the way, always believing in me. Without them and their support this would not have been possible.

This thesis was carried out under the project PTDC/MAR-EST/3048/2014, financed by Foundation for Science and Technology (FCT).

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

Resumo

Os avanços tecnológicos de instrumentos altamente sensíveis permitem a descoberta de várias substâncias e compostos que anteriormente não eram detectados. Muitos destes compostos não são completamente eliminados pelas estações de tratamento de efluentes. Como tal, podem espalhar-se diretamente pelo ambiente ou através de efluentes de águas tratadas e não-tratadas, contaminando rios, estuários e interagindo com organismos não-alvo. Apesar de as suas concentrações no ambiente aquático serem geralmente baixas (ng/L até μg/L), estes compostos são desenhados de modo a produzir efeitos biológicos a baixas concentrações, em humanos e organismos alvo, atuando em vias metabólicas específicas. Nas últimas décadas tem havido uma preocupação com os efeitos da exposição a fármacos em organismos aquáticos e, como tal, tem havido um crescente esforço para monitorizar e avaliar a sua presença e impactos no ambiente aquático. Embora a literatura acerca da ocorrência e impactos de fármacos no ambiente aquático esteja a crescer, ainda existem lacunas significativas no conhecimento desta temática.

Este trabalho procura fornecer, de forma quantitativa, conhecimento acerca dos efeitos de antidepressivos na reprodução de peixes e crustáceos através de uma meta-análise. Visto que os estudos em Biologia medem vários desfechos, neste caso várias medidas reprodutivas, é de esperar que haja dependência entre elas. Esta dependência pode vir, por exemplo, de grupos controlo partilhados, efeitos diferentes medidos nos mesmos indivíduos ou estudos que partilham a mesma espécie de organismo, entre outras. Deste modo, foi empregue uma meta-análise multi-nível que tem em conta a dependência entre os tamanhos dos efeitos obtidos (*effect sizes*), ou seja, permite que vários tamanhos dos efeitos venham do mesmo estudo, e considera que a variância se encontra distribuída por três componentes: variância amostral dos tamanhos dos efeitos extraídos (nível 1), variância intra-estudo (nível 2) e variância entre-estudos (nível 3).

Foi feita uma pesquisa sistemática da literatura na base de dados Web of Knowledge, usando palavras-chave relativas: à classe de fármacos utilizados e compostos específicos cujos efeitos já foram reportados; à classe de organismos e às respostas reprodutivas, desde o ano 2000 a 2019. Os estudos foram incluídos de acordo com critérios específicos de seleção. Entre eles: exposição ao fármaco era nos organismos (não em culturas de células); exposição na água (organismos não eram injetados com o fármaco); o estudo descrevia a espécie, os contaminantes e as concentrações testadas; o estudo descrevia as respostas específicas a serem medidas e indicava a duração da exposição. Dos 49 artigos obtidos, após remoção de duplicados, apenas 19 estudos cumpriram os critérios estipulados pelo que a análise consistiu em 19 estudos de literatura que investigavam os efeitos da exposição a antidepressivos na reprodução de peixes e crustáceos.

Devido à enorme variedade de respostas reprodutivas medidas nos estudos, estas foram agrupados em cinco

classes — Fecundidade, Eclodibilidade, Respostas Moleculares, Comportamento Reprodutivo e Maturação Sexual.

Os tamanhos dos efeitos foram estimados a partir de uma medida proposta por Hedges, o *Hedges' g*, visto ser uma estimativa não-enviesada da métrica usual para os tamanhos dos efeitos, o *Cohen's d*. De modo a evitar potenciais *outliers*, foram removidos tamanhos dos efeitos superiores a 3.29 ou inferiores a -3.29. Seguindo os procedimentos padrão em estudos meta-analíticos foi ajustado um modelo de ordenada na origem (sem moderadores) aos dados. Posteriormente, foi efetuada uma análise de heterogeneidade (i.e., variação extra não causada pelo acaso), através de testes log-verosimilhança que comparam o ajuste de um modelo com a componente de variância que queremos testar, contra outro modelo que carece desta mesma componente. Deste modo, é possivel determinar se é necessário ter em conta esta componente de variância. Adicionalmente, foram criados gráficos da distribuição da variância total por cada um dos níveis considerados pelo modelo e foi aplicada também a regra dos 75% para avaliar a heterogeneidade, como forma complementar aos testes log-verosimilhança.

Por forma a determinar se a variação adicional podia ser explicada por outros fatores, foi também testado o efeito moderador das variáveis concentração (medida em mg/L), tempo de exposição (em dias), grupo de organismos e fármaco, incluindo-as separadamente no modelo. Os preditores contínuos foram padronizados de modo a obter estimativas comparáveis entre estudos. A padronização permite também observar a magnitude das diferenças em número de desvios padrões. A análise dos moderadores não pode ser testada adicionando variáveis ao modelo e comparando com o ajuste de outro modelo sem as mesmas variáveis, visto estarmos a usar o método de estimação de máxima verosimilhança restrita (REML). Assim, para avaliar o possivel efeito moderador das variáveis, foi efetuado um teste "*omnibus*", baseado na distribuição F, no qual os preditores são incluídos no modelo e os coeficientes de cada categoria da variável (no caso de ser categórica) são, sob H_0, iguais a 0. Caso um deles seja significativamente diferente de 0, considera-se que há um efeito moderador significativo da variável em questão.

Adicionalmente, foi também efectuada uma análise ao viés de publicação (*Publication Bias*) recorrendo a gráficos funil (funnel plots) e ao teste de Egger.

Os resultados mostraram que, no geral, não existe associação significativa entre a exposição a antidepressivos e a reprodução de peixes e crustáceos, ao contrário do que indicam as revisões de literatura clássicas. Contudo, a análise de moderadores mostrou que a concentração é um moderador significativo para a fecundidade de peixes e crustáceos. Quando esta variável é incorporada no modelo, peixes e crustáceos mostraram resultados contrários. O efeito obtido para os primeiros é pequeno, embora positivo ($g = 0.368$), indicando que um aumento na concentração de exposição resulta num aumento de fecundidade para peixes, enquanto que os últimos apresentaram um efeito negativo, de magnitude um pouco maior, ($g = -0.453$) que aponta para fecundidade reduzida em crustáceos com um aumento na concentração de exposição. Após a incorporação da variável concentração no modelo relativo aos dados de fecundidade, a variância entre-estudos sofreu uma redução de aproximadamente 79%, o que mostra que o modelo que tem em contra a concentração de exposição apresenta informação mais precisa que o modelo sem moderadores. Quanto à análise do viés de publicação, esta indicou que a distribuição assimétrica dos efeitos (para os dados de fecundidade com peixes e crustáceos juntos, e para fecundidade considerando apenas crustáceos) pode dever-se a outros fatores que

não o viés de publicação, já que o gráfico funil apresentava efeitos "em falta" maioritariamente em áreas de elevada significância estatística.

Geralmente, as revisões narrativas reportam efeitos significativos porque os estudos incluídos, reportam efeitos significativos individualmente. Neste trabalho, 15 dos 19 estudos incluídos apresentam resultados significativos. Contudo quando comparamos os estudos através de uma análise quantitativa, com respostas padronizadas, o mesmo não se verifica, dado que esta não suporta a ideia de que há efeitos gerais na reprodução. Algumas limitações deste estudo são relativas aos dados, com a maioria dos estudos a focar-se apenas num composto (fluoxetina) e apenas em espécies de água doce.

Após a análise de moderadores ainda se verificou a existência de heterogeneidade significativa por explicar. Esta poderá ser explicada por outros fatores que podem eventualmente influenciar a associação entre exposição a antidepressivos e efeitos na reprodução, tais como a temperatura de exposição, pH ou salinidade, entre outros. Contudo a falta de informação relativa a estes possíveis moderadores inviabilizou que o seu efeito pudesse ser testado.

Os resultados deste trabalho apontam no sentido de que embora os antidepressivos tenham efeitos nas respostas reprodutivas de organismos aquáticos, como as revisões narrativas indicam, outros fatores podem ter um papel importante. Não obstante, são necessários mais estudos nesta área visto que as meta-análises são estatisticamente mais potentes, quantos mais estudos incluirem.

Palavras-Chave: Fármacos, Antidepressivos, Organismos aquáticos, Efeitos na reprodução, Meta-análise.

Abstract

Pharmaceutical compounds have been discharged into the environment for as long as they have been produced. Many of these pharmaceuticals can disperse directly into the environment or through sewage plants, contaminating rivers, estuaries and interacting with non-target organisms. Though their concentrations in the aquatic environment are often low, these drugs are designed to produce biological effects at low concentrations, targeting specific metabolic pathways. Though research on the impacts of pharmaceuticals in aquatic organisms is growing substantially, we still lack clear understanding on the topic.

This study aims to provide quantitative insight on the effects of antidepressants in the reproduction of fish and crustaceans. To address this, multilevel meta-analysis was performed. It consisted of 19 studies from research literature investigating effects of antidepressant exposure in fish or crustaceans' reproduction according to specific selection criteria. Following standard procedures used in meta-analytic studies, heterogeneity analysis was performed and the moderating effect of concentration, exposure time, organism group and toxicant was tested. Additionally, publication bias was also addressed.

The results showed that, overall, there is no significant association between antidepressant exposure and the reproduction of fish and crustaceans. Moderator analysis revealed, however, that concentration is a significant moderator for fish and crustatea fecundity. Fish and crustatea showed contrary results when accounting for this moderator. The former had a small, yet positive effect ($g = 0.368$), meaning that increased concentrations result in increased fish fecundity, and the latter a negative, in magnitude slightly larger, effect ($g = -0.453$), meaning a decrease in crustaceans' fecundity with a concentration rise. These findings help us understand that though antidepressants can have effects on reproductive outcomes in aquatic organisms, as narrative reviews address, other factors can have an important role. In spite of this, more research on the subject is need since meta-analysis are only as statistically powerful as the number of studies they include.

Keywords: Pharmaceuticals, Antidepressants, Aquatic organisms, Effects on reproduction, Meta-analysis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This study provides an overview on the research of the effects of pharmaceuticals on aquatic organisms as well as a syntheses on previous findings on this topic. The data for this study is therefore a compilation of published studies and their reported measures. To produce a combined analysis of these findings a meta-analytic approach will be used. Moderator analysis will also be performed in order to examine whether some factors influence the effects of pharmaceuticals' exposure, along with proposing some models that explain the effects of exposure on reproductive endpoints and testing them using hierarchical meta-analysis.

This chapter contains the motivation for this study. It introduces the issue and states the objectives, summarizing information on pharmaceuticals and their presence in the environment and briefly describes the structure of the thesis.

1.1 Motivation for the Study

1.1.1 Pharmaceuticals: What are They?

Pharmaceuticals are chemical substances used to diagnose, mitigate, treat, cure or prevent diseases. They have been present in our lives since the 1850's and allow us to live in our modern way. The pharmaceutical industry of today has had major development in comparison to it's roots as these compounds were first obtained through natural sources, but are now mostly synthesized in the laboratory.

There are many types of pharmaceuticals and they can be classified according to their active ingredient or their mode of action (MOA), having various therapeutic purposes (e.g., antibiotic, analgesic, antidepressant, etc...) (Kümmerer, 2010). These compounds are usually designed to target very specific biological pathways at extremely low concentrations and there has been a substantial increase of their usage in medical and veterinary fields, in the early 2000's for example, with around 3000 different substances in use in the European Union for human medical purposes (Fent et al., 2006). In general, the low volatility (i.e., the polar nature) of pharmaceuticals will prevent their escape from the aquatic realm, which means that distribution in the

environment will occur mainly through aqueous transport and also via food chain dispersal (Daughton and Ternes, 1999).

These features make pharmaceuticals, and other personal care products, emerging pollutants of priority concern and their presence in the environment should be assessed for evaluation of possible effects on aquatic life. As mentioned in EU Directive 2008/105/EC (European Parliament, 2008), EU Directive 2013/39/EU (European Commission, 2013) and EU Decision 2015/495 (The European Commission, 2015).

1.1.2 Presence in the Environment

The occurrence of pharmaceuticals and personal care products in the environment has been increasingly detected in the last years because they are inevitably discharged into the environment. They are ubiquitous contaminants in the aquatic environment, with adverse biological effects linked to exposure, and thus have been classified as contaminants of emerging concern (Kümmerer, 2010). The term "emerging" does not necessarily mean that the presence of these compounds in the environment is recent, but rather that there has been a development of new techniques to detect and quantify them or that there is a recent concern in it's increasing concentration and environmental effects (Glassmeyer et al., 2007).

Pharmaceuticals and personal care products are continuously released into the aquatic environment through multiple dispersal pathways. Treated and un-treated wastewater effluents are the primary source of contamination of aquatic systems, as some of these compounds and their metabolites are not completely eliminated by sewage treatment plants (Fent et al., 2006; Glassmeyer et al., 2007; Fatta-Kassinos et al., 2011). The removal of micropollutants depends on their physical-chemical properties (e.g. hydrophobicity and biodegradability), treatment conditions and other factors. Currently, the percentage of removal of micropollutants from wastewater treatment plants ranges from 18.8% to 91.1%, making this a variable and incomplete method that should be optimized in order to create an impenetrable barrier to micropollutants (Yunlong et al., 2014).

Ingestion and excretion of these chemicals by humans are generally followed by disposal via wastewater. So municipal and household wastes are the main human route to the environment. Wastewater from hospitals, not surprisingly, has a higher concentration of pharmaceuticals than municipal wastewater. However, their usage compared to the general public is lower, thus their share of the total load is also lower (Schuster et al., 2008).

Coastal and transition areas are of added concern, since they are interfaces of sea trade and industrial activities, holiday destinations, are largely occupied by human settlements and are also areas of high biological importance. They not only drain land effluent discharges, but it is estimated that 23% of the world's population lives within 100 km distance of the coast, which includes estuaries, where population densities are about three times higher than the global world average (Nicholls et al., 2007). Also, according to Seto et al. (2012), the five biodiversity hotspots that are forecasted to have their largest percentage of land areas become urban occupy coastal regions or are islands.

Moreover, activities like shipping, specifically cruises and large passenger ships can cause many environmental problems because of the critical emissions of exhaust gases and wastewater. This type of industry has

been quickly growing due to increased tourism, having more than 20 million passengers every year. Though regulated, their wastewater discharges lack effective treatment and assessment with some micropollutants present at concentrations up to 100 μg/L which could be harmful for aquatic life. For instance, the estimated annual load of Ibuprofen for a ship with 4000 people exceeds 3.3Kg (Westhof et al., 2016).

When it comes to veterinary applications via aquaculture, land based animal husbandry or livestock productions, pharmaceuticals are used for disease prevention, treatment of animals, growth promoters, among other purposes (antibiotics and azoles are the most common). For example, Reis-Santos et al. (2018) found various pharmaceuticals in surface waters of Tejo estuary (Portugal), including from veterinary use, albeit antibiotics, β-blockers, antihypertensives and anti-inflammatories were the most frequently detected ($> 90\%$). After their intake, some pharmaceuticals may be excreted as metabolites but only to a certain extent. For instance, a study by Kümmerer (2003) indicated that 75% of the antibiotics used in Germany are excreted unchanged, which means they are still active. These metabolites may have a lower effect on organisms than the parent compounds with the exception of prodrugs. Nałecz-Jawecki (2007) demonstrated that norfluoxetin was 50% more toxic than its precursor (Fluoxetine) in their essays. Following excretion from animals, manure application to fields may lead to pharmaceuticals entering the aquatic systems via runoff or drainage (Fent et al., 2006).

Regarding the pharmaceutical industry, production facilities may be the most important point sources in specific locations, and they contain by far the highest levels of pharmaceuticals in any effluent, with concentration values reaching tens of mg/L (Larsson et al., 2007).

The fate of pharmaceuticals in the environment depends on several environmental conditions (salinity, pH, turbidity, light penetration, oxidation, etc) and their own physical and chemical properties (Glassmeyer et al., 2007). As these compounds reach the environment some suffer structural changes through a variety of biotic and abiotic processes. Bacteria and fungi contribute to alter pharmaceutical compounds mostly in surface-, ground- and seawater while non-biological processes include: hydrolysis, photolysis, adsorption, biodegradation, amongst others (Kümmerer, 2010). Additionally, sorption is another factor that might regulate their presence, since binding to other particles or complexes can cause a loss in activity and/or detectability. A study by Maskaoui and Zhou (2010) provided direct evidence that sorption to colloids is an important sink for pharmaceuticals in the aquatic environment. Such strong pharmaceutical/colloid interactions may provide a long-term storage of pharmaceuticals, hence, increasing their persistence while reducing their bioavailability in the environment.

1.1.3 Exposure and Effects to Non-target Organisms

It has been shown that the range of pharmaceuticals' concentrations in surface water and sewage treatment plants is relatively low (ranging from ng/L to μg/L), and increased awareness of their occurrence has grown with the development of new techniques able to determine polar compounds at trace quantities (Fent et al., 2006; Kümmerer, 2010). Despite this, it is known that pharmaceuticals have a high biological specificity and are designed to produce biological effects at low concentrations. Moreover, many biological and metabolic pathways are in most cases evolutionary conserved (Furuhagen et al., 2014). As such, these compounds could

potentially lead to point source acute toxicity and chronic exposure of non-target organisms at different levels of biological organization (Huerta et al., 2012; Crane et al., 2006).

According to Fonseca and Reis-Santos (2018), considering coastal and marine environments, the most frequent group of organisms in pharmaceutical accumulation and toxicity studies are mollusks (69 studies), followed by crustaceans (32 studies) and fish (27 studies). When it comes to fish, several effects have been reported following pharmaceutical exposure. For instance, behavioural effects can differ between species and intake via food can be an important exposure route since on average 46% of the pharmaceuticals are ingested with prey and end up accumulating in the predator, *Perca fluviatilis* (Brodin et al., 2014). Some estrogens like 17-α-ethinylestradiol (EE_2) at concentrations as low as ng L^{-1} can alter sexual differentiation in males favoring the sex-ratio towards females (Länge et al., 2001).

Antidepressants like Fluoxetine, Fluvoxamine or Setraline, which are serotonin reuptake inhibitors (SSRIs), are widely used. Serotonin is a neurotransmitter that participates in essential regulatory and endocrine functions related to neuronal and hormonal mechanisms. These antidepressants act by blocking the reuptake of serotonin from the pre-synaptic nerve cleft and their presence in the environment could obviously lead to adverse effects to aquatic life (Fent et al., 2006).

There are various reports of antidepressants' effects on organisms' reproduction. For instance, in the western mosquitofish (*Gambusia affinis*), after exposure from juvenile to adult lifestages to 71μg/L of Fluoxetine, there was a delay in development of sexual morphology features in both males and females, showing that chronic exposure can lead to a delay in sexual development, but only in concentrations 3 to 4 times higher than those previously found in the environment (Henry and Black, 2008).

Additionally, some effects on fecundity and molecular responses include: Lister et al. (2009), in which Zebrafish exposed to 32 μg/L Fluoxetine for 7 days exhibited a significant reduction in egg production and a fall in ovarian estradiol concentration. Female goldfish exposed to 0.54 μg/L of Fluoxetine for 14 days exhibited a significant decrease in circulating estradiol (E_2) (Mennigen et al., 2017). Furthermore, Venlafaxine exposure in zebrafish decreased, though not significantly, the 11-ketotestosterone plasma concentration at low (0.5 μg/L) and high (10 μg/L) concentrations (Higgins et al., 2013). Some studies, however, had different findings. For instance, Parrott and Metcalfe (2017) found that fathead minnows females exposed to Venlafaxine at 88 μg/L produced 46% more eggs than control and that environmentally relevant concentrations of this compound (0.8 μg/L and 8.8 μg/L) did not cause any adverse effects in this species.

Effects on reproductive behaviour of fish following exposure to antidepressants have also been reported. A study by Weinberger and Klaper (2014) reported that concentrations of 1 μg/L of Fluoxetine significantly altered nest building and defensive behaviours in fathead minnows (*Pimephales promelas*) male fish, depending on exposure time. This exposure also limited female egg production due to significant male aggressive behaviour, which led to the death of females in the first two weeks of exposure.

On other aquatic organisms, the effects are also manifold. Antidepressants have been shown to affect spawning ability and larval release and also disrupt locomotion and reduce fecundity in snails. Crustaceans are also affected, for example, in Flaherty and Dodson (2005), chronic exposure to Fluoxetine (36 μg/L) significantly increased egg production in *Daphnia magna*. Campos (2012) reproductive assays of *Daphnia*

magna showed that exposure to SSRIs increased juvenile development rate, clutch size, and there was also a significant increase in offspring number for females exposed to Fluvoxamine (7 μg/L and 30 μg/L) or Fluoxetine (40 μg/L and 80 μg/L), albeit they also tested levels of food rations. Also, when exposed to 30 μg/L of Fluoxetine for 21 days, *Daphnia magna* individuals showed a slightly higher number of offspring compared to the control, though not significant (Varano et al., 2017). However, Campos et al. (2016) showed that at low food conditions, Fluoxetine exposure (40 μg/L) causes significant increase in offspring numbers for *Daphnia magna* individuals.

These examples, along with many others show that pharmaceuticals clearly have adverse effects in aquatic organisms at concentrations currently present in the environment. But ultimately, though literature on occurrence and impacts of these contaminants in the environment is steadily growing, there is still a lack of information that we should aim to tackle.

1.1.4 Increasing Research Effort

The vast majority of studies investigating occurrence and impacts of pharmaceuticals on the aquatic environment encompass freshwater systems. This is probably due to the assumption that dispersion and dilution processes, in coastal and marine areas, would suffice to lessen or cancel any potential effects. In their review Fong and Ford (2014) discuss the effects of antidepressants' exposure on molluscs and crustaceans highlighting impacts on locomotion, growth, behaviour, metabolism and reproduction, however the majority of information referred to freshwater invertebrates. When it comes to fish, information on toxicity studies of antidepressants focuses mainly in freshwater species as well (Corcoran et al., 2010; Fonseca and Reis-Santos, 2018). Reported data for marine organisms is also limited to other classes of pharmaceuticals, with antibiotics being the most commonly tested in aquaculture studies (Gaw et al., 2014). Despite this, knowledge on coastal and marine biota is steadily growing with various reports showcasing the presence of pharmaceuticals in transition and marine environments at concentrations potentially adverse to different levels of biological complexity (Du et al., 2016; Aminot et al., 2016; Gaw et al., 2014). Fonseca and Reis-Santos (2018) found 124 studies focusing on ecotoxicology of pharmaceuticals in coastal and marine biota as well as on bioaccumulation in wild coastal and marine organisms, which shows a clear increase from the 49 studies found by Gaw et al. (2014).

For marine risk assessment, a prioritization approach of both generic and novel prescription pharmaceuticals is recommended by Gaw et al. (2014), which suggested the use of measures like predicted exposure concentration (PEC_{marine}) and predicted no-effect concentrations ($PNEC_{marine}$). Thought for the former, the mode of action (MOA) of the pharmaceutical should be considered by evaluating Adverse Outcome Pathways — link between exposure, the interaction of a contaminant at the molecular level within a cell and an adverse outcome — in freshwater organisms and extrapolating this information to marine species (Gaw et al., 2014). There is also a need for the development of data sharing mechanisms on pharmaceutical occurrence and their effects in non-target species. Daughton (2014) proposed the creation of a database on the occurrence of APIs (Active Pharmaceutical Ingredients) in the environment, which would allow for a real-time perspective on what pharmaceuticals have escaped our attention.

1.2 Main Objectives

In order to evaluate the effects of exposure to pharmaceuticals in aquatic organisms' reproductive success this study will employ a meta-analysis, which is a set of statistical procedures that allow us to combine results of primary studies and draw overall conclusions of a specific topic (Crombie and Davies, 2009). The main objective of this work is to identify and to assess major patterns of effects of pharmaceuticals' exposure on different reproductive endpoints of aquatic organisms, through a hierarchical meta-analysis approach. It will focus on one class of pharmaceuticals, antidepressants, which have been reported to affect aquatic organisms at concentrations currently found in the environment (Arnold et al., 2014).

1.3 Structure of the Thesis

First, in Chapter 2 a brief description of meta-analysis as a method will be introduced, with essential background information on the procedure applied to conduct it, from literature search and inclusion criteria to the model construction and moderator analysis. Chapter 3 reports the results of this analysis. Starting with study characteristics, followed by the actual meta-analytic results, respective heterogeneity analysis, moderator analysis and assessment of publication bias. Finally, in Chapter 4 the results obtained are discussed in light of other reported findings, limitations are discussed as well as some insight on future research of pharmaceutical exposure studies and lastly the conclusion for this study is presented. The Appendix contains a table with the list of articles included in this meta-analysis, boxplots of the distribution of concentrations used in each study and of the distribution of exposure time, the forest plots obtained for each subset of data, as well as the variance distribution plots - representing the percentage of total variance over the three levels of the meta-analytic model - and the funnel plots for assessment of publication bias.

Chapter 2

Methodology

This chapter will first provide background information on the methodology for this study. It starts by introducing meta-analysis as a method, comparing it to the traditional narrative review approach, and it describes some of its applications in contemporary research fields. The initial steps for the meta-analysis process will be described, in particular, literature search and inclusion criteria. Effect sizes considered and publication bias will also be addressed here. Lastly, it will cover everything to do with the model itself, from correlation of effect sizes to the moderator analysis.

2.1 Background - What is a Meta-analysis and Why Use It

Meta-analysis and systematic reviews emerged as a quantitative way of summarizing the results of the reviewed studies and combining them to obtain an overall effect (or effect size). To search published studies and decide which ones are to be included, there is a set of criteria defined in advance. These criteria are chosen by researchers, which means there is some level of subjectivity attached to them, therefore this method is not completely objective (Borenstein et al., 2009). Even so, because the decisions are clearly specified, the mechanisms are transparent. Unlike the traditional approach (the narrative review), where reviewers assign some level of importance to each study, meta-analysis uses mathematical formulas to assign the weights to each study (Borenstein et al., 2009).

The narrative review has several disadvantages. It has no mechanism to synthesize the p-values of the different studies, thus it cannot give us a measure of the magnitude of the effect nor discuss it's significance, which is one of the main questions posed by a study. It is more prone to bias since statistically significant studies are more likely to be included in the review, and the informal synthesis might be biased by prior beliefs of the reviewer. It is also insufficient when synthesizing findings from contradictory results, especially for a large number of studies (Hunter et al., 1991). As such, the systematic review and subsequent statistical analysis provides an objective and reproducible way to report our results (Pigott, 2012). Furthermore, meta-analysis can be seen as an extension of the formulas used in primary studies (Borenstein et al., 2009). For example, in a primary study we might report a mean and standard deviation for the subjects while, similarly,

in meta-analysis we might report a mean and standard deviation for a treatment effect (Crombie and Davies, 2009).

Many research fields nowadays resort to meta-analysis and systematic reviews to draw more accurate conclusions when synthesizing data regarding the effects of some intervention. In medicine, reviews look at interventions in various areas of healthcare including surgery, drugs, acupuncture and social interventions (Borenstein et al., 2009). Additionally, they are used to assess epidemiological associations between diseases and exposure factors and evaluate the performance of diagnostic tests (Sutton and Higgins, 2008). In the pharmaceutical industry, many studies are conducted to test the efficacy of a drug and meta-analysis are then applied to obtain more precise estimates of the drug's effect (Lièvre et al., 2002). Education has also been influenced by meta-analytic studies, which encompass, for instance, analyzing how student performance is influenced by teacher's credentials, the relation between distance and traditional classroom learning, among others Bernard et al. (2004); Falchikov and Goldfinch (2000). Other applications include psychology in assessment of personality change, aggressive behaviours due to media violence and obviously to compare and select treatments for psychological disorders like obsessive-compulsive disorder, impulsivity disorder, bulimia nervosa, depression, phobias, and panic disorder (Edens et al., 2007). Criminology research has also used meta analysis to test the effectiveness of different programs in reducing criminal behaviour (Wilson, 2001). In business, to optimize the hiring process based on validity tests and to compare for example, different programs on the employee training procedure (Hausknecht et al., 2004). Lastly, meta-analysis are being used in ecology to determine environmental impacts caused by exotic species, climate change, contaminants and also to optimize conservation interventions (Borenstein et al., 2009).

Usually, we can divide the meta-analytic process in the following steps: (1) formulating the research problem; (2) systematic search of relevant studies; (3) data extraction from the primary studies; (4) statistical analysis; and (5) interpretation and overall conclusions. These steps will be addressed in the next section.

2.2 Procedure

2.2.1 Literature Search and Inclusion Criteria

As any other meta-analysis, the process started with the research question, in this case, to know if pharmaceuticals have an overall effect on the reproduction of aquatic organisms and how these effects can vary when considering different reproductive endpoints. Following this, selection criteria were established and literature search was conducted. The search was undertaken in the Web of Knowledge^{TM} database (November 2018), which was the first citation database for scholarly literature and research articles (Bakkalbasi et al., 2006). Due to the massive study field, i.e., ecotoxicological data and available studies on pharmaceuticals' effects on aquatic life cover many classes of organisms, various endpoints and several classes of pharmaceuticals, the scope of this study was limited to antidepressants' effects on reproduction, which have been reported to affect aquatic organisms at concentrations currently found in the environment (Arnold et al., 2014). Additionally, the search ranged from the years 2000 to 2019 and only fish and crustaceans were considered (following a brief representation analysis on available studies).

The search was conducted using key-words regarding: (1) class of pharmaceutical and specific compounds known to produce effects; (2) class of organism and (3) reproduction related endpoints. Two separate searches were done, one focusing only on fish and another focusing on crustaceans. Thus, the following strings were used:

1. (Antidepressant* OR Fluoxetine OR Venlafaxine OR Citalopram OR Fluvoxamine OR Mianserin OR Norfluoxetine OR Paroxetine OR Sertraline) AND Fish AND (reprod* OR fecund* OR sex*)
2. (Antidepressant* OR Fluoxetine OR Venlafaxine OR Citalopram OR Fluvoxamine OR Mianserin OR Norfluoxetine OR Paroxetine OR Sertraline) AND Crustac* AND (reprod* OR fecund* OR sex*)

The following PRISMA diagram (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) describes the steps in the systematic search and provides details on the obtained results, thus facilitating transparency and reproducibility.

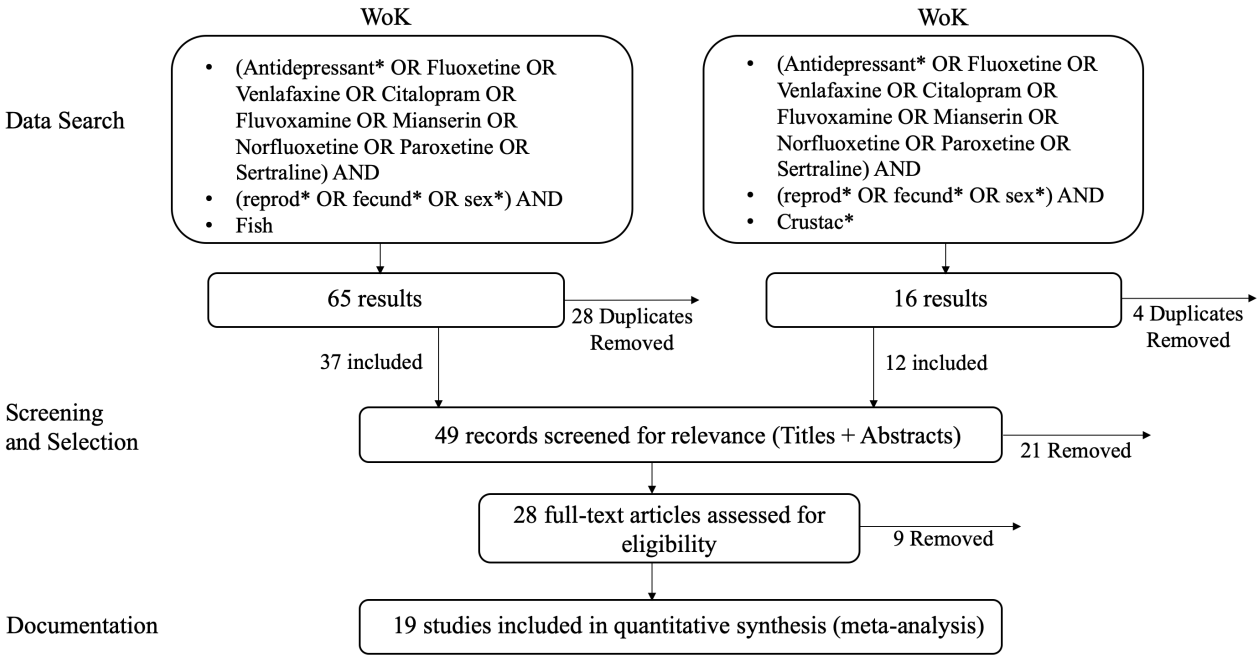


Figure 2.1: PRISMA diagram. WoK - Web of Knowledge

The first search produced 65 results from which only 37 were kept. The second search only produced 16 results from which 12 were pre-selected. In the end, a total of 49 entries were kept including 6 review articles.

The criteria for inclusion of a study were: (1) exposure was on full organisms; (2) organisms were subject to waterborne exposure and were not injected with the pharma; (3) the study clearly described the species, the individual contaminants used and the concentrations tested; (4) described the specific endpoints measured and (5) stated the duration of exposure.

Following these criteria the studies were then screened for relevance through a search in the title and abstract using words like: injec, vitro, cell line, intraperitoneal or mixtures. This selection resulted in 28 articles.

Afterwards, the resulting articles were assessed in their integrity for eligibility using the same criteria. Nine articles were discarded due to data not being extractable (e.g. not reporting on sample size or variation response), leaving the analysis with 19 viable studies. The next step was data extraction regarding reproduction related endpoints along with extra information like pH and temperature of the experimental design, lifestage of the organism and its sex, whenever available.

Due to the vast amount of different responses measured, these were grouped into five classes of endpoints. Each class included the following measures:

1. **Fecundity** - Total eggs; Eggs per female; Offspring size and number; Cluch size; (among others).
2. **Hatchability** - Eggs hatched; Hatch-success
3. **Molecular Responses** - Hormones like Estrogen (E2), Testosterone, Luteinizing hormone (LH), Growth hormone (GH), Follicle-stimulating hormone (FSH), Vasotocin, Isotocin, 11-KT (Ketotestosterone); Aromatase; Vasa-marker; Vitelogenin.
4. **Reproductive Behaviour** - Courting events; Courting displays; Nesting tile behaviours like cleaning or number of visits.
5. **Sexual Maturation** - Gonadosomatic index (GSI); Tubercule index; Genital pappilae; Ovipositor area; Sum of secundary sex characteristics; Black spot; Gonopodium; Age at first reproduction.

Data was then compiled into an Excel document.

2.2.2 Effect Sizes

The measures of interest are expressed as effect sizes, which are defined by Cohen (1988) as *the degree to which [a] phenomenon is present in the population*. With the null hypothesis (H_0) stating, in this case, that there is no difference between the exposed group and the control group. The larger the effect size, the larger the probability of H_0 being false. One commonly used metric for the effect size, and the one this study will use is the standardized mean difference. Since raw mean differences are only useful if all the primary studies use the same scale of measurement, which is almost never guaranteed in biological studies, we divide the mean differences by the study's standard deviation to create an index that is comparable across studies (Borenstein et al., 2009), i.e. the standardized mean difference, δ. Using standard deviations allows us to obtain a quantitative measure on the performance of the intervention group when compared to the control group, by number of standard deviations, or a proportion of standard deviations .

The sample estimate of this measure is often designated by Cohen's d, and because the two groups under comparison (experimental and control) are independent and assuming homogeneity of the population variances, it can be estimated by

$$d = \frac{\overline{X}_1 - \overline{X}_2}{S_{within}} \tag{2.1}$$

where $\overline{X}_1$ and $\overline{X}_2$ are the means of the two independent groups, in this study, the experimental and control groups, respectively, and S_{within} is the pooled standard deviation,

$$S_{within} = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \tag{2.2}$$

where n_1 and n_2 are the sample sizes in the two groups, and S_1 and S_2 are the respective standard deviations.

The reason variances are pooled is so that we obtain a more accurate estimate of their common value as estimates may differ substantially in the two groups, even if we assume that the population standard deviations are the same, i.e, $\sigma_1 = \sigma_2 = \sigma$.

The approximate variance for d and it's standard error are estimated by:

$$V_d = \frac{n_1+n_2}{n_1 n_2} + \frac{d^2}{2(n_1+n_2)} \tag{2.3}$$

and

$$SE_d = \sqrt{V_d} \tag{2.4}$$

respectively.

One particular aspect about Cohen's d is that it overestimates the absolute value of the population parameter δ in the case of small samples. Because of this, the current study will use an unbiased estimate of Cohen's d proposed by Hedges (2008) commonly called Hedges' g. There is a need for a correction factor (J) to convert Cohen's d into Hedges' g and it can be obtained by the following approximation,

$$J = 1 - \frac{3}{4df - 1} \tag{2.5}$$

where df corresponds to the degrees of freedom used to estimate S_{within}, which in the case of two independent groups is $n_1 + n_2 - 2$. From here, it follows that

$$g = J \times d \tag{2.6}$$

$$V_g = J^2 \times V_d \tag{2.7}$$

and

$$SE_g = \sqrt{V_g} \tag{2.8}$$

The correction factor (J) will be less than 1 and so the value of g is always smaller than d. Some authors even ignore the term J^2 from the variance of g because it is usually close to unity. Other authors consider different expressions for the variance but in practice they will be almost identical, unless sample sizes are too small.

The equations here presented provide context for the effect size measure to be used. However, it is up to the researcher to choose the quantitative measure that is most fitting for their work, and whilst the differences between Cohen's d and Hedges' g might be small, the correction factor is recommended (Borenstein et al., 2009). Thus, this study will use Hedges' g as the effect size metric.

To verify if there were any outlying effect sizes, the data was screened for standardized z-values larger than 3.29 or smaller than -3.29 as recommended by Tabachnick and Fidell (2014). Z-values are interpreted the same way as Hedges' g values, both representing the number of standard deviations in which the two groups differ.

2.2.3 Reporting Meta-analysis Findings

Forest plots are graphical representations of the effect sizes and have become the conventional way of reporting the results of meta-analysis as they combine the results and display them in a set of axis to which the eye is more drawn to. Each line represents an effect size from a study, indicated by the black boxes. The size of the boxes indicates the relative weight of that effect size in the analysis. With the three-level modeling approach still being a random-effects model, the relative weights are assigned in a more balanced way so that larger studies do not overweight the smaller studies.

Putting it simply:

- The larger the sample size, the shorter the horizontal line and the bigger the black box representing the point estimate. For these effect sizes it is less likely that they will cross the null effect line because the 95% confidence intervals will have a smaller range.
- The smaller the sample size, the wider the horizontal line and the smaller the black box representing the point estimate. In this case it is more likely those effect sizes will cross the line of null effect (as the 95% confidence intervals will be much wider).

At the bottom of the graph, the diamond shape represents de overall estimate and confidence interval of the combined studies.

When a moderator variable is added to the model the overall effect is no longer visible at the bottom of the graph. However, gray diamond shapes are added to each line representing the effect size of that study when study-level covariates (moderators) are included. The forest plots here presented have the following information specified: in the left corner is shown the author of the study, publication year, endpoint measured, pharmaceutical tested and its concentration in mg/L; whilst in the right corner, the effect sizes and their respective 95% confidence intervals are depicted.

2.2.4 Publication Bias

When performing a systematic review or a meta-analysis, one must take into consideration that some studies might have escaped our criteria in the selection process. These missing studies might be a *random* subset of all the relevant studies, or they might be *systematically* different than the ones we located. If the former is true, then failing to include these studies will lead to loss of information, wider confidence intervals, and a lower statistical power, but will have no systematic impact on the effect size. However, if the latter is true, our sample will be biased (Borenstein et al., 2009). Studies reporting larger effect sizes are more likely to have statistically significant results than the ones reporting smaller effect sizes, and consequently are more likely to be published, given any sample size (Rothstein et al., 2006). This leads to publication bias, as the unpublished studies with lower effect sizes were not included, and thus our estimate of the pooled effect might be higher than the true effect size.

Meta-analysts could assume that research is published regardless of statistical significance and that authors do not selectively report their results. As this is very unlikely, publication bias should be assessed. According to Borenstein et al. (2009), we expect the bias to increase as the sample size decreases:

- Because they involve large commitments of time and resources, large studies are likely to be published, whether the results are significant or not;
- Studies with a moderate sample size will generally have significant effects, though these are at risk of missing;
- Small studies are at greatest risk of being lost, as their sample sizes are smaller, only the largest effects are likely to be significant, unlike the small and moderate ones which end up unpublished.

Considering these assumptions, the most commonly used methods for assessing publication bias are the following:

Fail-safe *N*: It tells us how many studies with mean value 0 do we need to add to the analysis before the cumulative effect would become statistically unsignificant (Long, 2001). There are two approaches for this method, Rosenthal's and Orwin's. The latter is preferred since it does not focus on statistical significance and allows us to determine how many studies would bring the overall effect to a level other than zero (Borenstein et al., 2009; Orwin, 2008). Using the small-threshold as criterion value for fail-safe studies, and zero for the mean effect size of the fail-safe studies the formula for calculating fail-safe N is as follows:

$$Fail\text{-}safe\ N = K\ \frac{r - 0.1}{0.1} \tag{2.9}$$

where K and r are the number of studies in the analysis and the mean effect size, respectively. The criterion value (the threshold for small effect) for fail-safe studies is 0.1.

Funnel Plot: Plots the effect size of the study (x-axis) against a measure of study size (y-axis). Sterne et al. (2001) notes that the use of the standard error as a measure of study size is generally a good choice since the statistical power of a trial is determined by both the sample size and the number of participants developing

the event of interest. Precision ($1/$standard error) can also be a recommended alternative (for the vertical axis) if there is an interest in emphasizing differences between larger studies.

Studies with a large sample size will cluster around the mean effect size and the smaller studies appear towards the bottom of the graph, and tend to be more spread out since they have more sampling error variation. If there is no publication bias, the studies will be symmetrically distributed around the pooled effect size, since the sampling error is random. In the case of publication bias, published studies (with large effect size) would stand towards the top of the "funnel", while the small and moderate studies would be missing as we reach the base.

Egger's Test: Provides a simple analysis of funnel plot asymmetry by performing a statistical regression test. Should the result be significant, we conclude that there is substantial asymmetry and that it could have been caused by publication bias (Egger et al., 1997). The null hypothesis being tested is $H_0 : \beta_0 = 0$ (no asymmetry). If there is funnel plot asymmetry, the regression line will not run through the origin, and thus β_0 provides a measure of asymmetry, while the slope β_1 represents the size and direction of effect (Sterne and Egger, 2005). The larger the deviation of β_0 from zero, the more pronounced the asymmetry.

***P*-Curve**: This approach to evaluate publication bias assumes that the bias does not occur due to unpublished non-significant results, but rather because they alter their data (e.g., selectively removing outliers, choosing different outcomes, controlling for different variables) in order for an unsignificant result to become significant. This is often called *p*-hacking and is widespread throughout science (Head et al., 2015). The *p*-curve is displayed in a graphic where the y-axis shows the percentage of test results and the x-axis shows the *p*-values. If there is no effect, the distribution of *p*-values should be uniform (a horizontal line). If there is an effect, the curve should be right-skewed, with a higher percentage of results with *p*-values in the range of 0.02 and 0.01. On the other hand, if there is a left-skewed curve this means that there is a much higher percentage of *p*-values in the range of 0.04 and 0.05. This indicates the researchers altered their data to get a significant result, which is usually the case at 0.05 *p*-values.

According to Simonsohn et al. (2014) the shape of the *p*-curve will vary in the following way:

- Sets of studies investigating effects that exist are expected to produce right-skewed *p*-curves.
- Sets of studies investigating effects that do not exist are expected to produce uniform *p*-curves.
- Sets of studies that are intensely *p*-hacked are expected to produce left-skewed *p*-curves.

The desirable shape is to have right-skewed *p*-curves, as only true effects are expected to generate them, they are diagnostic of evidential value.

This study will employ the funnel plot and Egger's test to assess publication bias since they complement each other, with the funnel plot being an intuitive visual display and Egger's test being available for multilevel-meta analytic models by extending the model to include the sampling standard errors (Sterne and Egger, 2005).

The funnel plot to be used is a contour-enhanced funnel plot which displays contour lines and areas for the following significance levels: $p < 0.05, < 0.025, < 0.01$. These plots assume a two-sided significance

test is performed in the individual studies. The advantage of this variant is that it allows us to distinguish publication bias from other causes of asymmetry. So, for example, if in areas of non-significance, studies appear to be missing, asymmetry could be indicative of publication bias. On the other hand, if the supposed missing studies are in areas of higher statistical significance, other factors might be causing asymmetry, such as variable study quality (Peters et al., 2008). Also to note that there might be causes other than the ones based on statistical significance, for publication bias (e.g., sample size and effect size) and researchers should have this into account when interpreting contour-enhanced funnel plots.

2.3 The Model

2.3.1 Correlation between Effect Sizes

Another important aspect to consider when conducting a meta-analysis is that traditional models (Fixed-effects and Random-effects) assume independence between effect sizes, which in biological meta-analysis is uncommon (see Vesterinen et al. (2014)). Correlation between effect sizes can come from multiple sources, such as shared study identity (e.g., effects sizes estimated from samples derived from the same population), shared measurements within a study (e.g., effect sizes derived from different traits measured in the same individuals), shared control group within study or shared species across studies (Nakagawa et al., 2017). These correlations might lead researchers to erroneous conclusions since the information is "inflated" and leads to overconfidence in the results (Van den Noortgate et al., 2013).

Previous solutions to dealing with this dependence included calculating a weighted or unweighted average of the dependent effect sizes, or selecting one effect size per study, among others. Cheung (2014) notes that not only these solutions limit statistical power, due to loss of information, but also the number of questions that can be explored via meta-analysis. Also, extracting one effect size per study implies that there is homogeneity of effect sizes, which may not be true.

Therefore, this study will consider a recent approach, a three-level meta-analytic model (Assink and Wibbelink, 2016), which takes into account interdependency of effect sizes (i.e., for example, they allow for multiple effect sizes to come from the same study) and considers that the variance is distributed over three levels: sampling variance of extracted effects sizes (Level 1); variance between effects sizes of the same study (within-study variance - Level 2); and between-study variance (Level 3), as illustrated by the following figure:

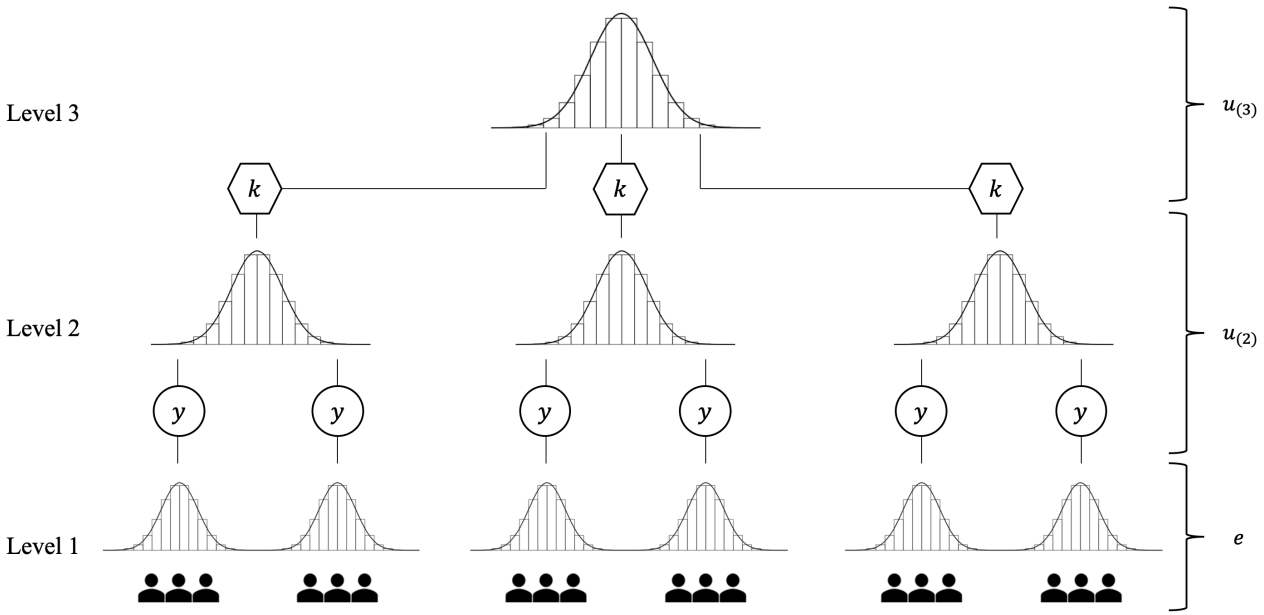


Figure 2.2: Three-level structure of the meta-analytic model; where y is the effect size, k is the study, e is the samplig error of extracted effect sizes, $u_{(2)}$ is the error relative to diferences between effects of the same study and $u_{(3)}$ is the error relative to diferences between studies - Adapted from: Harrer and Ebert (2018)

Accordingly, the data will have a hierarchical structure, with effect sizes in level 2 nested within studies at level 3, whilst level 1 refers to the sampling variance of effect sizes. The model can be defined by the following mathematical formulae (Cheung, 2014):

Level 1 model:

$$y_{ij} = \lambda_{ij} + e_{ij} \tag{2.10}$$

Level 2 model:

$$\lambda_{ij} = k_j + u_{(2)ij} \tag{2.11}$$

Level 3 model:

$$k_j = \beta_0 + u_{(3)j} \tag{2.12}$$

Where λ_{ij} is the "true" effect size, y_{ij} its estimator in the ith effect size in cluster j. As an example, y_{ij} might represent one of multiple effect sizes in the jth study, or it might represent one of the studies in the jth cultural group of a cross-cultural meta-analysis. In this case, the former is true. Additionally, k_j is the average effect size in the jth cluster, β_0 is the average population effect and the sampling error is given by e_{ij}. $Var(u_{(2)ij}) = \tau^2{}_{(2)}$ and $Var(u_{(3)j}) = \tau^2{}_{(3)}$ are the study-specific level 2 and level 3 heterogeneity, respectively. These equations can be combined into one single equation as defined by Cheung (2014):

$$y_{ij} = \beta_0 + u_{(2)ij} + u_{(3)j} + e_{ij} \tag{2.13}$$

2.3.2 Heterogeneity

To properly interpret the overall effect obtained in the meta-analysis, one must assess the heterogeneity among effect sizes. Cochran's Q is a common test statistic used to evaluate the existence of effect size heterogeneity. It allows us to test whether between-study variance $Var(u_{(3)})$ (or τ^2 as it is usually designated) is non-zero, i.e. $H_0 : \tau^2 = 0$. The mathematical formula for Cochran's Q (adapted from: Chen and Peace (2013)) is:

$$Q = \sum_{i=1}^{K} w_i(\hat{\delta}_i - \hat{\delta})^2$$

where:

- w_i is the weight from the ith study given by the inverse-variance
- $\hat{\delta}_i$ is the ith study effect size
- $\hat{\delta}$ is the summary effect
- K is the number of studies included

The $\hat{\tau}^2$ index is the variance of the true effect sizes. We cannot calculate this variance straightforward because we do not know what the true effect sizes are, but using observed data it can be estimated as follows:

$$\hat{\tau}^2 = \frac{Q - (K-1)}{U} \tag{2.14}$$

which is the DerSimonian-Laird method of moments for τ^2 (DerSimonian and Laird, 1986), with

$$U = \sum_{i=1}^{K} w_i - \frac{\sum_{i=1}^{K} {w_i}^2}{\sum_{i=1}^{K} w_i} \tag{2.15}$$

Notably the estimated variance $\hat{\tau}^2$ can be less than zero even though the true variance of τ^2 can never be. This can happen if the sampling error leads to $Q < K-1$, but when this occurs the estimated $\hat{\tau}^2$ is set to zero. The $\hat{\tau}^2$ is also, as mentioned, an estimate of the between-study variance in the analysis of the true effects (Chen and Peace, 2013).

The statistical power of the Q statistic depends on the number of studies. If the number of studies is small the power is low and if the number of studies is large, the power is high (Huedo-Medina et al., 2006). To solve the problems of the Q statistic and the non comparability of the between-studies variance, τ^2, among meta-analyses with different effect-size metrics, the I^2 index was proposed by Higgins (2003). This measure quantifies the percentage of excess variance to total variance, and can be estimated as follows:

$$I^2 = \frac{Q - (K-1)}{Q} \times 100\% \tag{2.16}$$

Its values can be interpreted as follows: a value of around 25% corresponds to low-heterogeneity, 50%, to moderate-heterogeneity and 75% to high-heterogeneity. Since there is a direct relationship between I^2 and τ^2 (Huedo-Medina et al., 2006) formalized by

$$I^2 = \frac{c\tau^2}{Q} \tag{2.17}$$

this measure is a good approach to indicate how much of the variance in effect sizes can be attributable to between-study variance.

To assess the significance of within-study variance (level 2) and between-study variance (level 3), a log-likelihood-ratio test will be performed for each. These tests will be executed by calling the `anova` function in `R`, which will compare the fit of the original model where the variance components are freely estimated against a model where one of these components is set to zero. To test the within-study variance component, the fit of the original model will be compared to the fit of a model where the variance at level 2 is manually fixed to zero, allowing us to determine if it is necessary to account for this variance component. As such, the null hypothesis will be $H_0 : \sigma^2(level2) = 0$ and the alternative hypothesis, $H_1 : \sigma^2(level2) > 0$. The same procedure will be used to test for between-study variance, with the corresponding hypothesis being: $H_0 : \sigma^2(level3) = 0$ *vs* $H_1 : \sigma^2(level3) > 0$.

In addition to this method, a different approach to examining heterogeneity is also recommended. If the data is comprised of very few studies and/or effect sizes, log-likelihood-ratio tests may not be significant despite having substantial within-study or between-study variance. This might lead researchers to avoid performing moderator analysis. Thus, a different perspective to examining heterogeneity is the 75% rule, which states that if less than 75% of the total variance is attributed to sampling variance (level 1) it can be assumed that there is substantial heterogeneity (Hunter et al., 1991).

2.3.3 Moderator Analysis

After quantifying the excess variation, i.e., variation not attributable to sampling error, it is important to understand the factors that might explain it. Moderators are the equivalent of predictors or explanatory variables in a linear regression and can explain high variation in the data (Nakagawa et al., 2017). In this meta-analysis the moderators to be tested are: the organism group (Fish or Crustacea); the toxicant, i.e., the chemical compound used (Bupropion, Citalopram, Fluoxetine, Fluvoxamine, Sertraline, Venlafaxine); exposure concentrations (expressed in mg/L) and exposure time (expressed in days). Continuous moderators were standardized, i.e. the mean was subtracted from each variable value and the result was divided by the standard deviation, using the `scale` function in `R`.

Model construction and analysis will be executed using `RStudio` (Version 1.1.456, RStudio Team (2015)), an interface of `R` software, with the help of the `rma.mv` function in the `metafor` package (Viechtbauer, 2015).

The method used for parameter estimation is the REstricted Maximum Likelihood (REML). It is recommended over other methods (Maximum-Likelihood, DerSimonian–Laird, Hunter–Schmidt) since it is asymp-

totically efficient and unbiased as discussed by Viechtbauer (2007). One downside of the REML method is that it does not allow us to compare a model with moderators with the one without the moderators through a log-likelihood-ratio test (Hox, 2010). As such, an *omnibus* test will be conducted to evaluate the significance of the effect of moderating variables, i.e., whether they should incorporate the model. Hence, the model used to test for the effect of moderators is referred as a meta-regression. The null hypothesis states that the coefficients associated to the moderators (i.e., betas) are equal to zero ($H_0 : \beta_1 = \ \beta_2 = ... = 0$) and will be tested against the alternative hypothesis ($\exists\, j_{\in\{1,...,m\}} : \beta_j \neq 0$). This test will be based on the F distribution with the degrees of freedom of the numerator being equal to the number of coefficients in the model and the degrees of freedom of the denominator being equal to $k - p$, with k representing the number of effect sizes and p the total number of coefficients in the model including the intercept.

The `rma.mv` function by default uses the normal distribution — Z distribution — in test statistics of individual coefficients and confidence intervals. However it as been shown that basing these statistics on the Z distribution may lead to an increase in the number of unjustified results (Ziegler et al., 2001). Therefore the calculations in this study take into account the Knapp and Hartung (2003) adjustment which accounts for this problem by basing the testing of individual coefficients on the t-distribution with $k - p$ degrees of freedom. Additionally, if the *omnibus* test is applied to an intercept-only model with a single moderator the value of the test statistic (F) will be equal to the square of the t value.

Chapter 3

Results

This chapter contains the results of the meta-analysis. First the results of the three-level meta-analytic model are reported, together with the respective heterogeneity analysis. Then the moderating effects of organism group, toxicant, concentration and exposure time are tested. Analyzing the combined effect sizes of every study as a whole would not make sense since they measure different endpoints. As such, a model was adjusted to the reproductive classes of endpoints separately (Fecundity, Molecular Responses, Reproductive behaviour and Sexual Maturation).

3.1 Literature Characteristics

The conducted search resulted in 19 valid and relevant research articles for analysis, from 2000 to 2019, with a total number of $k = 298$ effect sizes. When screening for outlying effect sizes, Fecundity data and Molecular responses data had Hedges' g values larger than 3.29 and lower than -3.29, as such, these were removed. The resulting total number of effect sizes was $k = 291$. The articles used in this meta-analysis are listed in Table A.1 (see appendix B, with descending alphabetical order of the reference). In addition to study references, the table contains the studied organism group, the endpoint classes the study encompasses and the DOI code. In Table 3.1 are displayed the number of articles for each subgroup and it is evident that many categories do not have any studies, even though one study can be present in more than one subgroup. There is a minimum number of studies to be considered when performing a meta-analysis, in this work models were only applied to categories that had a minimum of $K \geq 3$ studies and $k \geq 20$ effect sizes. Taking this into account, the Hatchability class and the Crustacea data relative to Sexual Maturation endpoints became redundant since they did not meet these criteria.

Table 3.1: Number of studies for each subgroup

	Endpoint Class				
Organism Group	Fecundity	Hatchability	Molecular Responses	Reproductive Behaviour	Sexual Maturation
Fish	3	2	7	6	6
Crustacea	6	0	0	0	2

3.2 Intercept-only Model

The syntax used to built the intercept-only model, taking fecundity data as an example, was the following:

```
fecund.overall <- rma.mv(y, v, random = list(~ 1 | effectsizeID, ~ 1 | studyID),
tdist=TRUE, data=fecund, method = "REML")

summary(fecund.overall, digits=3)
```

Taking a closer look at the arguments for this syntax:

- `fecund.overall =` the name of the object in which the the results of this function will be stored. Since it is estimating an overall effect for fecundity data, this was the name chosen;
- `y =` variable containing all effect sizes, expressed in Hedges' g;
- `v =` variable containing all sampling variances (as estimated in Equation 2.3);
- `random =` argument that specifies the model has a random-effects structure and since the primary studies are assumed to be a random sample of the population of studies, the random-effects model is adequate;
- `list(~ 1 | effectsizeID, ~ 1 | studyID) =` element that defines the three-level structure of the model. The random- effects variance is denoted by `~ 1` and is assigned to a grouping variable by "`|`". The second level of the model is defined by `effectsizeID` which contains unique identifiers for all effect sizes and the third level at which the variance between effect sizes is distributed is defined by `studyID` which contains specific identifiers for all the studies in the data set. To note that the sampling variance is assumed to be known and therefore is not specified in the model.
- `tdist=TRUE =` argument specifying the test statistics and confidence intervals are based on the t-distribution;
- `data=fecund =` the object containing the name of the data set;
- `method="REML" =` method used for estimating the model parameters.

The results of fitting an intercept-only model to the data can be seen on Table 3.2 and were obtained by running the sintax above for each set of data regarding different reproductive endpoints. The effect size

estimates (Hedges' g values) are negative except for Sexual Maturation endpoints, which means that pharmaceutical exposure has an apparent positive effect on: (1) Sexual Maturation of fish and crustacea, and an apparent negative effect on: (1) Fecundity of fish and crustacea; (2) Molecular Responses and Reproductive Behaviour of fish.

Globally, none of the models were significant, which means that there is no overall significant association between pharmaceutical exposure and the reproduction of aquatic organisms, specifically fish and crustaceans. Furthermore, all of Hedges' g estimates here presented are regarded as small, according to the criteria formulated by Cohen (1988), which states that differences of 0.2, 0.5 and 0.8 standard deviations are considered as small, medium and large effect sizes, respectively.

Table 3.2: Results of the three-level intercept-only model

Endpoint Class			**Model Results**				**Test for Heterogeneity**	
					95% conf. int.			
	k	g	t	p	low	high	Q	p
Fecundity	73	-0.028	-0.210	0.834	-0.290	0.235	144.525	<.001**
Fish	26	-0.134	-0.396	0.696	-0.832	0.564	25.431	0.438
Crustacea	47	-0.042	-0.233	0.817	-0.402	0.318	119.079	<.001**
Hatchability								
Fish	-	-	-	-	-	-	-	-
Molecular Responses								
Fish	110	-0.201	-1.231	0.221	-0.524	0.122	210.563	<.001**
Reproductive Behaviour								
Fish	42	-0.065	-0.774	0.443	-0.234	0.104	32.913	0.812
Sexual Maturation	56	0.033	0.288	0.775	-0.197	0.264	81.920	0.011*
Fish	52	0.052	0.385	0.702	-0.219	0.323	79.313	0.007**
Crustacea	-	-	-	-	-	-	-	-

Notes: Fish and Crustacea refer to the subgroups; k = number of effect sizes; g = Hedges' g estimate; t = t -test statistic; Q = Chochran's Q test statistic; p = p-values; Significance codes: 0.01'**', 0.05 '*'

3.3 Heterogeneity Analysis

As depicted in Table 3.3 there is significant within-study variance in: (1) Fecundity data (as a whole and when subgrouped into crustacea) and (2) Molecular Responses data (only studied in fish). There is also significant between-study variance in: (1) Molecular Responses data. This means that for these categories the variability in effect sizes is bigger than what would be expected from sampling variance alone. In addition to this, the distribution of the total variance over the three levels of the meta-analytic model was determined using formulas given by Cheung (2014). In Appendix D the percentage of total variance attributed to each level is shown for every subgroup of data analyzed. Making use of the 75% rule stated by Hunter et al. (1991), those with sampling variance (level 1) less than 75% of the total variance have substantial heterogeneity.

Therefore it makes sense to perform moderator analysis to see if some variability can be explained by study and/or effect size characteristics.

Table 3.3: Results of the log-likelihood-ratio tests

	Heterogeneity Analysis					
	Within-Study Variance			Between-Study Variance		
Endpoint Class	AIC Full	AIC Reduced	*p*	AIC Full	AIC Reduced	*p*
Fecundity	167.363	175.300	0.002*	167.363	168.900	0.060
Fish	48.947	46.947	1.000	48.947	47.505	0.455
Crustacea	117.527	130.063	<.001**	117.527	118.716	0.074
Hatchability						
Fish	-	-	-	-	-	-
Molecular Responses						
Fish	254.033	265.872	<.001**	254.033	261.995	0.002*
Reproductive Behaviour						
Fish	39.874	37.874	1.000	39.874	38.530	0.418
Sexual maturation	119.274	118.397	0.289	119.274	119.182	0.167
Fish	114.650	114.199	0.213	114.650	114.685	0.154
Crustacea	-	-	-	-	-	-

Notes: *p* = *p*-values; Degrees of freedom: Full model = 3, Reduced model = 2;
Significance codes: 0.01 '**', 0.05 '*'

3.4 Moderator Analysis

Here are presented the results of the *omnibus* tests performed for each subgroup of data in order to determine whether there is a potential moderating effect of some variables, in this case, the organism group, toxicant, exposure concentrations and exposure time.

These results will be shown separately, for discrete and continuous moderators.

3.4.1 Categorical Moderators

Organism Group

Table 3.4 shows the results of the moderating effect of Organism Group, which is a study-level moderator. Both Fecundity and Sexual Maturation were the only classes in which there were studies for both groups, fish and crustacea, hence, the only tested. Fecundity data showed a borderline *p*-value for the significance of between-study variance ($p = 0.06$). Complementing this test with the 75% rule and looking at Figure D.1, the sampling variance is bellow 75%, thus substantial heterogeneity can be assumed. Even though Sexual

Maturation data did not show significant within-study and between-study variance from the significance tests, the variance distribution plots (Appendix D) show a sampling variance below 75% of the total variance and thus, this class was tested for moderating effect as well.

For this moderator, it was tested if effect sizes from studies with fish were significantly different from studies with crustaceans. One dummy variable named `Fish` (coded as 1 for studies with fish or 0 if not) was created and was then added to the model by writing `mods = ~Fish`. Therefore, considering the different endpoint classes:

Fecundity:

The results of the *omnibus* test reveal that the overall effect is not moderated by the organism group ($F(1,71) = 0.000$; $p = 0.997$). When the organism group is considered as a moderating variable, the mean effect of the studies with crustaceans (reference category) is -0.038 and not significantly different from zero ($t = -0.217$; $p = 0.829$). The mean effect of the studies with fish is $-0.001 + (-0.038) = -0.039$, so there is an apparent negative effect of pharmaceutical exposure to the fecundity of fish and crustacea, when accounting for the variability of this moderator. This mean effect is, however, not significant ($t = -0.003$; $p = 0.997$) which is in line with the results from the *omnibus* test, which uses the F-distribution since the Knapp & Hartung adjustment was applied (Knapp and Hartung, 2003)

To note that since only one potential moderating variable is being tested in this model, the F-test statistic value will be the square of the t-test statistic. For Fecundity data, $F = 0.000$ is approximately the square of $t = -0.003$.

Sexual Maturation:

The overall effect is not moderated by the organism group ($F(1,54) = 0.103$; $p = 0.75$). The mean effect of the studies with fish is equal to $-0.055 + 0.106 = 0.051$ and is not significantly different from the mean effect of the reference category ($g = -0.055$) as the *omnibus* test results indicate. Although not significant, there appears to be a positive effect of pharmaceutical exposure on the sexual maturation of fish and a negative effect on crustaceans.

For both Fecundity and Sexual Maturation data, the test for heterogeneity reveals that there is still significant unexplained variance in the model, when the moderating effect of the organism group is accounted for.

Even though this variable did not have a significant moderating effect, the data analysis was still subdivided for the remaining moderators to see if there were any differences in the effects of pharmaceutical exposure for both groups.

Table 3.4: Results of the Organism Group moderator model

Organism Group										
Endpoint Class			**Model Results**						**Test for Heterogeneity**	
					95% conf. int.					
	k	*g*	*t*	*p*	low	high	*F*	*p*	*Q*	*p*
Fecundity										
Crustacea (Reference)	73	-0.038	-0.217	0.829	-0.382	0.307	0.000	0.997	144.509	<.001**
Fish		-0.001	-0.003	0.997	-0.623	0.620				
Sexual Maturation										
Crustacea (Reference)	56	-0.055	-0.181	0.857	-0.666	0.556	0.103	0.750	81.840	<.001**
Fish		0.106	0.320	0.750	-0.559	0.772				

Notes: Crustacea and Fish refer to the moderator categories; k = number of effect sizes; g = Hedges' g estimate (regression coefficient); $t = t$ -test statistic for the regression coefficient; $F = F$-test statistic (*omnibus* test); Q = Chochran's Q test statistic; $p = p$-values; Significance codes: 0.01 '**', 0.05 '*'

Toxicant

In this section it was examined whether the overall association between pharmaceutical exposure and its effects on fish and crustacean reproduction is moderated by the type of compound used, i.e., the toxicant. Five compounds were used in the included studies: Bupropion, Citalopram, Fluoxetine, Fluvoxamine, Sertraline and Venlafaxine. Though some of them were not part of the inclusion criteria for the literature search they are still viable for this study. Bupropion, for example, does not have the same effects as SSRIs (Selective serotonin reuptake inhibitors) or SNRIs (Serotonin–norepinephrine reuptake inhibitors), but can be classified as a norepinephrine/dopamine-reuptake inhibitor (NDRI). Nonetheless, it is still an antidepressant.

Since Fluoxetine is the most common used pharmaceutical in studies that test antidepressant exposure, being present in all the classes of reproductive endpoints, this was the category chosen to be the reference category. `R` chooses the reference category by alphabetical order, as such, the reference category was changed making use of the `relevel` function with `ref = "fluoxetine"`.

In this case, as there were many different categories for each subset of data, no dummy variables were created. If Venlafaxine was not tested in Reproductive Behaviour studies, for example, then this predictor would be dropped automatically from the model seeing that it is redundant. So, this moderator was included in the model by adding the syntax `mods = ~Toxicant`.

The results for the Toxicant moderator are presented in Table 3.5.

Fecundity:

From the model results we derive that there is no moderating effect of Toxicant as the results of the *omnibus* test point out ($F(3,69) = 0.639;\ p = 0.593$ for Fecundity data; $F(1,24) = 0.461;\ p = 0.503$ for fish Fecundity data and $F(3,43) = 0.639;\ p = 0.593$ for crustacea Fecundity data). The mean effect of Fluoxetine is equal to -0.117 and does not significantly deviate from zero, since $t(69) = -0.718,\ p = 0.475$. None of the categories were significantly different from the reference category. When compared to Fluoxetine, Ven-

lafaxine has a positive effect on the Fecundity of fish and crustaceans, contrary to the other antidepressants, though none are significant.

Considering fish data, only Fluoxetine and Venlafaxine were tested. The former has a mean effect of -0.455 which is not significantly different from zero ($t(24) = -0.747,\ p = 0.463$). And the latter has a mean effect equal to $0.683 + (-0.455) = 0.228$ which is not significantly different from the reference category ($t(24) = 0.679,\ p = 0.503$) which is in line with the results of the *omnibus* test. There is a trend that the effect of Venlafaxine is once again apparently positive in the Fecundity of fish.

In crustacean data, the mean effect of Fluoxetine is -0.092 and not significantly different from zero ($t(43) = -0.418,\ p = 0.678$). None of the remaining categories are significantly different from Fluoxetine and there appears to be a negative effect of Fluvoxamine, Sertraline, and Venlafaxine when compared to Fluoxetine, and a trending positive effect of Venlafaxine when compared to Fluoxetine.

Molecular Responses and Reproductive Behaviour:

Toxicant does not have a moderating effect on the molecular responses and reproductive behaviour of fish as the *omnibus* tests indicate (Table 3.5). All the compounds seem to have an apparent negative mean effect.

In Molecular Responses data, estimated effect sizes are regarded as a high effect size for Bupropion ($g = -0.762 + (-0.044) = -0.806$), and moderate effect sizes for Sertraline ($g = -0.669 + (-0.044) = -0.713$) and Venlafaxine ($g = -0.719 + (-0.044) = -0.763$) (Cohen, 1988), with none of them being significantly different from Fluoxetine. In Reproductive Behaviour, exposure to both Fluoxetine ($g = -0.064$) and Citalopram ($g = -0.039 + (-0.064) = -0.103$) point to a negative effect of the Reproductive Behaviour of fish.

Sexual Maturation:

Considering fish and crustacea data together, it is not evident that Toxicant is a moderator variable for Sexual Maturation since the results of the *omnibus* test are not significant ($F(4,51) = 1.99;\ p = 0.110$). Bupropion has a negative effect equal to $-0.84 + (-0.198) = -1.038$ and is significantly different from the reference class ($t(51) = -2.183;\ p = 0.034$), as opposed to the remaining categories, i.e., Fluvoxamine, Sertraline and Venlafaxine.

The results are similar for fish data alone since there is still no moderating effect of Toxicant ($F(3,48) = 2.062;\ p = 0.118$). Bupropion has a negative effect ($g = -0.838 + (-0.196) = -1.034$) and is significantly different from Fluoxetine ($t(48) = -2.062;\ p = 0.045$), which has a positive effect ($g = 0.196$). The remaining categories seem to have negative effects on sexual maturation and none are significantly different from the reference category.

In sum, we can conclude that the overall association between pharmaceutical exposure and its effects on Sexual Maturation of fish and crustacea is not moderated by the toxicant/antidepressant tested in studies thus far.

Table 3.5: Results of the Toxicant moderator model

Toxicant										
Endpoint		**Model Results**							**Test for Heterogeneity**	
					95 conf interval					
	k	*g*	*t*	*p*	low	high	*F*	*p*	*Q*	*p*
Fecundity										
Fluoxetine (Reference)	73	-0.117	-0.718	0.475	-0.444	0.209	0.639	0.593	138.580	<.001**
Fluvoxamine		0.086	0.360	0.720	-0.390	0.561				
Sertraline		-0.054	-0.111	0.912	-1.014	0.907				
Venlafaxine		0.417	1.216	0.228	-0.267	1.101				
Fish										
Fluoxetine (Reference)	26	-0.455	-0.747	0.463	-1.715	0.804	0.461	0.503	24.009	0.461
Venlafaxine		0.683	0.679	0.503	-1.393	2.759				
Crustacea										
Fluoxetine (Reference)	47	-0.092	-0.418	0.678	-0.537	0.353	0.316	0.814	114.031	<.001**
Fluvoxamine		0.067	0.240	0.812	-0.494	0.627				
Sertraline		-0.046	-0.077	0.939	-1.259	1.166				
Venlafaxine		0.483	0.78	0.44	-0.765	1.730				
Hatchability										
Fish	-	-	-	-	-	-	-	-	-	-
Molecular Responses										
Fish										
Fluoxetine (Reference)	110	-0.044	-0.248	0.804	-0.399	0.310	1.650	0.182	196.838	<.001**
Bupropion		-0.762	-1.507	0.135	-1.765	0.241				
Sertraline		-0.669	-1.328	0.187	-1.668	0.330				
Venlafaxine		-0.719	-1.896	0.061	-1.47	0.033				
Reproductive Behaviour										
Fish										
Fluoxetine (Reference)	42	-0.064	-0.469	0.642	-0.339	0.211	0.029	0.865	32.150	0.807
Citalopram		-0.039	-0.171	0.865	-0.499	0.422				
Sexual maturation										
Fluoxetine (Reference)	56	0.198	1.779	0.081	-0.025	0.421	1.990	0.110	70.982	0.034*
Bupropion		-0.840	-2.183	0.034*	-1.612	-0.067				
Fluvoxamine		-0.537	-1.370	0.177	-1.325	0.250				
Sertraline		-0.444	-1.160	0.252	-1.214	0.325				
Venlafaxine		-0.308	-1.633	0.109	-0.687	0.071				
Fish										
Fluoxetine (Reference)	52	0.196	1.621	0.112	-0.047	0.439	2.062	0.118	70.098	0.020*
Bupropion		-0.838	-2.062	0.045*	-1.655	-0.021				
Sertraline		-0.445	-1.098	0.277	-1.259	0.370				
Venlafaxine		-0.306	-1.537	0.131	-0.707	0.094				
Crustacea	-	-	-	-	-	-	-	-	-	-

Notes: k = number of effect sizes; g = Hedges' g estimate (regression coefficient); t = t -test statistic for the regression coefficient; F = F-test statistic (*omnibus* test); Q = Chochran's Q test statistic; p = p-values; Significance codes: 0.01 '**', 0.05 '*'

3.4.2 Continuous Moderators

Concentration

Testing for the moderating effect of concentration is of interest since changes in the amount of drug per units of volume means changes in the amount of active pharmaceutical ingredient present in the water, and no matter how small these changes are, they may have an effect on the organism's reproduction. Since the F-test value is the square of the t-test statistic the results of the t-test were here omitted.

Looking at the mean effects separately:

Fecundity:

From Table 3.6 we can derive that Concentration is a significant moderator for Fecundity data (as a whole, and when fish and crustaceans are considered separetly), since the *omnibus* test results are significant ($F(1,71) = 10.329$; $p = 0.002$ for Fecundity data; $F(1,24) = 7.859$; $p = 0.01$ for fish Fecundity data; $F(1,45) = 16.198$; $p < 0.001$ for crustaceans Fecundity data).

The mean effect of pharmaceutical exposure is equal to -0.285, i.e., the higher the concentration of the compound, the lower the fecundity of fish and crustaceans. Since this predictor is standardized, another way to look at these findings is: for each fold difference in the standard deviation — $sd(Concentration) = 0.0826$ — that the concentration deviates from the mean concentration, there is a decrease of 0.285 standard deviations in the fecundity of exposed fish and crustaceans, when data on both groups is considered together.

However, when considering fish and crustacea data separately, with an increase in concentration, pharmaceutical exposure has a positive effect on the fecundity of fish ($g = 0.368$) and a negative effect on the fecundity of crustaceans ($g = -0.453$), both being significant.

The value of the intercept is not interpreted here, since it represents the mean (effect) of effect sizes extracted from studies in which the concentration used was the mean concentration. Hence, like in regular regression analysis, it does not represent the mean effect of the reference category.

Molecular Responses and Reproductive Behaviour:

Concentration is not a significant moderator when considering either molecular responses or reproductive behaviour (Table 3.6).

Sexual Maturation:

Sexual Maturation data regarding crustacea only had two studies, therefore there are two groups of results. One that includes both fish and crustacea (though only the two studies), and the other with studies on fish only.

Concentration is not a significant moderator ($F(1,54) = 0.134$; $p = 0.715$ for Sexual Maturation data; $F(1,50) = 0.009$; $p = 0.925$ for fish Sexual Maturation Data).

The test for heterogeneity reveals that when accounting for the moderating effect of Concentration there is

still significant unexplained variance except for fish Fecundity data ($Q = 17.112;\ p = 0.844$) and Reproductive Behaviour data ($Q = 32.886;\ p = 0.780$).

In sum, we can conclude that Concentration is a significant moderator of the overall association between pharmaceutical exposure and its effects on the fecundity of fish and crustaceans.

Table 3.6: Results of the Concentration moderator model

Concentration									
Endpoint Class		**Model Results**						**Test for Heterogeneity**	
				95% conf. int.					
	k	*g*	se	low	high	*F*	*p*	*Q*	*p*
Fecundity	73	-0.285	0.089	-0.463	-0.108	10.329	0.002**	127.608	<.001**
Fish	26	0.368	0.131	0.097	0.639	7.859	0.010**	17.112	0.844
Crust	47	-0.453	0.113	-0.680	-0.227	16.198	<.001**	94.991	<.001**
Hatchability									
Fish	-	-	-	-	-	-	-	-	-
Molecular Responses									
Fish	110	-0.083	0.072	-0.225	0.059	1.336	0.250	210.116	<.001**
Reproductive Behaviour									
Fish	42	-0.012	0.058	-0.129	0.104	0.047	0.830	32.886	0.780
Sexual Maturation	56	0.035	0.095	-0.155	0.224	0.134	0.715	80.649	0.011*
Fish	52	0.010	0.103	-0.197	0.216	0.009	0.925	78.463	<.001**
Crustacea	-	-	-	-	-	-	-	-	-

Notes: Crustacea and Fish refer to the subgroups; *k* = number of effect sizes; *g* = Hedges' *g* estimate (regression coefficient); se = standard error; *F* = *F*-test statistic (*omnibus* test); *Q* = Chochran's *Q* test statistic; *p* = *p*-values;
Significance codes: 0.01'**', 0.05 '*'

Exposure Time

It is also interesting to test for the moderating effect of the exposure time since changes over time may influence the strength of association between pharmaceutical exposure and its effect on reproduction.

Table 3.7 shows the results of the model considering Exposure Time.

Fecundity:

Exposure Time is not a significant moderator for every subgroup as the *omnibus* test results indicate. When Exposure Time is present in the model the mean effect size is positive for Fecundity data as a whole ($g = 0.117$) and for fish only ($g = 0.005$), but it is negative for crustaceans ($g = -0.009$).

Molecular Responses and Reproductive Behaviour:

There is no moderating effect of Exposure time when considering either Molecular responses or Reproductive Behaviour. For both classes the effects are negative and considered small.

Sexual Maturation:

Exposure Time is not a significant moderator for Sexual Maturation data as a whole ($F(1,54) = 0.087$; $p = 0.769$, nor for fish data only ($F(1,50) = 0.039$; $p = 0.844$). The regression coefficient is positive (though not significant) for Sexual Maturation data ($g = 0.037$), which could suggest a trend, where a higher exposure time favors the development of sexual maturation characteristics, but also age at first reproduction of fish and crustacea. The mean effect for fish data is zero, which means that there is no effect of the exposure time on the Sexual maturation of fish.

Overall, we can conclude that Exposure Time is not a significant moderator of the effects of pharmaceutical exposure on the reproduction of fish and crustaceans.

Table 3.7: Results of the Exposure Time moderator model

Exposure Time									
Endpoint Class		**Model Results**						**Test for Heterogeneity**	
				95% conf. int.					
	k	*g*	se	low	high	*F*	*p*	*Q*	*p*
Fecundity	73	0.117	0.154	-0.190	0.424	0.574	0.451	143.599	<.001**
Fish	26	0.005	0.006	-0.009	0.018	0.536	0.471	23.908	0.165
Crust	47	-0.009	0.030	-0.070	0.052	0.086	0.770	118.338	<.001**
Hatchability									
Fish	-	-	-	-	-	-	-	-	-
Molecular Responses									
Fish	110	-0.117	0.132	-0.379	0.145	0.786	0.377	206.917	<.001**
Reproductive Behaviour									
Fish	42	-0.017	0.110	-0.239	0.204	0.025	0.875	32.676	0.788
Sexual Maturation	56	0.037	0.126	-0.215	0.290	0.087	0.769	81.642	<.001**
Fish	52	0.000	0.002	-0.004	0.005	0.039	0.844	79.098	<.001**
Crustacea	-	-	-	-	-	-	-	-	-

Notes: Crustacea and Fish refer to the subgroups; k = number of effect sizes; g = Hedges' g estimate (regression coefficient); se = standard error; F = F-test statistic (*omnibus* test); Q = Chochran's Q test statistic; p = p-values; Significance codes: 0.01 '**', 0.05 '*'

Generally, the next step would be to create a multiple moderator model in which all the significant moderators were included and test the significance of the residual within-study and between-study variance. Since Concentration was the only variable that had a significant moderating effect, in the model for Fecundity data, it does not make sense to perform the analysis of multiple moderators.

3.5 Publication Bias

Appendix E contains the funnel plots for every subset of data and the results of their respective Egger's test, which was obtained by modifying the meta-analytic model to include the standard error of the effect

sizes as a moderator variable. The overall relationship between sample size and effect size is considered asymmetrical if the intercept of the regression test significantly deviates from zero (Sterne and Egger, 2005).

We found that, Fecundity data, fish and crustaceans together, and Fecundity data considering only crustaceans have significant asymmetry of the funnel plots (Figure E.1, Figure E.3), which could suggest publication bias. Hatchability data too, has significant asymmetry but it is not here considered, as it was not fit for model analysis since it only had two studies.

Funnel plot asymmetry could mean publication bias, but as Sterne et al. (2001) reported, asymmetry could also mean *small-study effects* which tell us that smaller studies are usually conducted less rigorously than larger ones and thus asymmetry might be the result of overestimation of effects. There was no significant asymmetry in the remaining funnel plots, so there is no evident bias for the published studies regarding the three other classes considered.

The problem with publication bias is that the estimated overall effect of our meta-analysis might be higher than the true effect size, since missing studies with lower effects were not considered due to the simple fact that they were never published.

Chapter 4

Discussion

The purpose of this meta-analysis was to provide an overview of quantitative studies on the effects of pharmaceuticals in the reproduction of aquatic organisms. While in classic literature reviews there might be a tendency to report that there are significant effects, due to focusing on significant effects by individual studies (for example in Fong and Ford (2014) and Santos et al. (2009)), in this meta-analysis we see that the variability in effects is so high that they might be specific for a certain species, or for a particular range of concentrations tested, or dependent on the studies' experimental design, among other things. Despite this, a meta-analysis may also be limited by various factors, first and foremost data quality.

In this chapter, the meta-analysis results will be discussed and compared to other individual findings and reviews, followed by a critical assessment of the limitations of this methodology and recommendations for future research.

4.1 Overall Effect and Moderating Effects

Does exposure to pharmaceuticals have repercussions in aquatic life reproduction? The present meta-analysis of 19 studies suggests that there are mostly negative effects, though none are significant, prior to moderator analysis. The average effect size for each endpoint class was: (1) -0.028 for Fecundity; (2) -0.201 for Molecular Responses; (3) -0.065 for Reproductive Behaviour; and (4) 0.033 for Sexual Maturation measures, which is a positive effect. A negative Hedges' g value indicates that the reproductive response is lower in the group exposed to pharmaceutical concentrations than in the control group, whereas a positive value indicates that the reproductive response is higher in the exposed group when compared to the control group.

The results of the test for heterogeneity, in the intercept-only model, revealed significant variation for all effect sizes since the p-values are smaller than 0.05, except for fish data regarding Fecundity endpoints and for Reproductive Behaviour endpoints, which were only available for fish. These results are, however, not very informative, as our interest does not lie within the variation between all effect sizes but rather in the within-study variance and between-study variance.

Combining the log-likelihood-ratio test results, which indicated there was significant variation between effect sizes within and across studies, and the 75% rule (Hunter et al., 1991) — Fecundity, Molecular Responses and Sexual Maturation data had sampling variance below 75%, thus indicating substantial heterogeneity, it made sense to perform moderator analysis to assess this excess in variability. When testing for moderators that explained the differences in variance over the three levels, the analysis revealed that Concentration was a significant moderator for the association between pharmaceutical exposure and its effects on the Fecundity of fish and crustaceans, when considered together and separately. With this predictor taken into account, the variance between-studies (level 3) decreased approximately 79% (from 0.0871 to 0.0182) for Fecundity data (fish and crustaceans), 60% (from 0.2648 to 0.1069) for Fecundity - fish data and 100% (from 0.1117 to ≈ 0) for Fecundity - crustacea data. This reduction can be seen in percentage points as well, in the variance distribution figures bellow, using Fecundity data considering fish and crustacea together as an example (Figure 4.1, Figure 4.2).

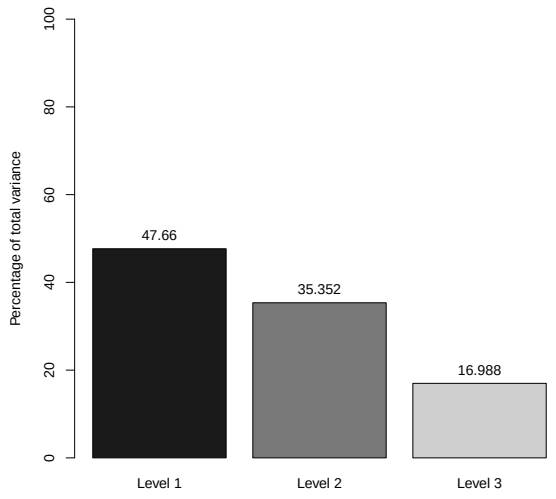


Figure 4.1: Variance distribution of Fecundity data

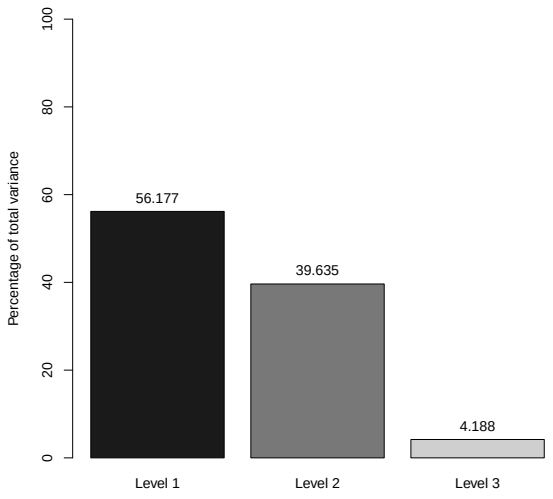


Figure 4.2: Variance distribution of Fecundity data with concentration taken into account

This means that $\approx 13\%$ of the total variance ($16.988\% - 4.188\%$), which was attributable to variance across studies (Level 3), could be explained by the difference in concentrations used between these studies. As such, the model with concentration as a moderator brings more accurate information than the intercept-only model. Moreover, Fecundity data regarding fish had $\approx 45\%$ between-study variance, i.e. attributable to level 3, and the remaining variance was distributed over level 1 (sampling variance), with 0% of the total variance attributable to within-study variance — Figure D.2. According to the 75% rule (Hunter et al., 1991), there should be substantial heterogeneity of effect sizes between studies, however the log-likelihood-ratio tests indicate this additional variation is not significant (Table 3.3), likely due to very few studies with low to unsubstantial within-study variability.

With concentration taken into account as moderating variable, the effect differed between fish and crus-

tacea in Fecundity data when considered separately, i.e., with an increase in concentration there was a positive effect of pharmaceuticals' exposure on the fecundity of fish, yet a negative effect on the fecundity of crustaceans. Or, from another point of view, for each fold difference in the standard deviation — $sd(Concentration - \text{fish data}) = 0.0286$; $sd(Concentration - \text{crustacea data}) = 0.0953$ — that the concentration deviates from the mean concentration, there is an increase of 0.368 standard deviations in the fecundity of exposed fish and a decrease of 0.453 standard deviations in the fecundity of exposed crustaceans. Furthermore, estimated coefficients are larger for Fecundity, which points out a stronger effect in Fecundity with the increase of concentration than in the other reproductive endpoint classes considered.

This difference between fish and crustaceans groups could be due to biological differences; data artifacts, since the collected sample of studies from fish and crustacea might not be representative of the whole population of species, or could be due to different species' responses to the toxicant used (Fluoxetine was the most common). These factors might also be responsible for the remaining significant unexplained heterogeneity between effect sizes — for Fecundity data together and considering just crustaceans — when the moderating effect of Concentration is accounted for (Test for Heterogeneity, Table 3.6). Parrott and Metcalfe (2017), for instance, found that fathead minnows exposed to Venlafaxine at 88 μg/L, the highest concentration tested, produced 46% more eggs per female than in the control group, with fish exposed throughout their life cycle from the embryo state to mating and reproductive stages. This supports the positive effect of concentration rise on the fecundity of fish obtained in this work. The overall Fecundity effect (with no predictor) being negative could mean that the crustacea data, when combined with fish data, was more representative (i.e. higher weight) to change the positive effect it had on fish separately, to a negative effect when both are considered. This is likely the case, keeping in mind that there were 47 effect sizes in crustacea studies and 26 effect sizes in fish studies.

Fluoxetine, the active ingredient in Prozac, is a commonly prescribed SSRI and widely used in pharmaceutical exposure studies. Out of 291 effect sizes from the global data set, 204 resulted from studies with organisms exposed to Fluoxetine. Because of this potential bias, the moderating effect of Toxicant was tested, yet it was not significant ($F(4,51) = 1.990$; $p = 0.110$). However, the antidepressant Bupropion was significantly different from Fluoxetine, due to data limitations, ($t = -2.183$; $p = 0.034$) in Sexual Maturation data. This result is not of particular interest to this study, but it serves to show that while there were only 2 effect sizes resulting from experimental designs that used Bupropion there were $k = 33$ out of 56 effect sizes respecting Fluoxetine.

4.2 Comparison with Individual Study Findings

Notably, the only crustacean organism model used in the studies included in the meta-analysis was *Daphnia magna*, so these results may be species specific. This clearly highlights data constraints/limitations.

When it comes to the effects on this species, several studies reported increased offspring production in *Daphnia magna* when exposed to Fluoxetine in a concentration range of tens of μg/L (Flaherty and Dodson, 2005; Campos et al., 2016; Campos, 2012; Varano et al., 2017). This points towards the existence of a positive effect of exposure on crustacean fecundity. However, these results are contrary to this meta-analysis' findings,

as the effect size estimated in this study points to a negative effect on crustacean's fecundity. Nonetheless, there is evidence that pharmaceuticals, more specifically antidepressants, affect aquatic life at concentrations currently found in the environment (ng/L range), including daphnid reproduction and development, as Fong and Ford (2014) address in their review. Santos et al. (2009) also reported that new data on sources, fate and effects of pharmaceuticals in the environment seem to indicate the possibility of a negative impact on different ecosystems.

Fluoxetine's global average concentrations in surface water are usually within the ng/L range, with a global review by aus der Beek et al. (2016) indicating a variation between 7 ng/L to 47 ng/L (see supplement 10 in aus der Beek et al. (2016)), while in this work the range of concentrations included was from 2.5 ng/L to 400 μg/L. Hence, some studies test exposure concentrations within the range present in the environment, yet the majority are well above. Additionally, in a study by Donnachie et al. (2016), prioritization of pharmaceuticals was investigated in terms of environmental risk, through an unbiased ranking approach. Fluoxetine was not only among the most cited, in the analyzed reviews, but was also ranked as a pharmaceutical of highest concern, however the authors note there could be some bias to literature-based assessments since the majority of toxicology studies tend to test the same compounds, to ensure results comparability to some degree.

In fish, chronic exposures to Fluoxetine have resulted in disruption of reproductive endpoints, including steroidogenesis and gametogenesis in both female and male fish (Mennigen et al., 2010). Other studies report effects on fecundity, e.g. decreased egg production (Lister et al., 2009), and molecular responses of various fish species, e.g. decreasing ovarian estradiol, plasma estradiol (Mennigen et al., 2017) or 11-ketotestosterone, (Higgins et al., 2013). These findings are according with the overall effect obtained in this study (albeit not statistically significant) as both Fecundity and Molecular Responses have a negative estimated effect size on fish, considering no predictors. Foran et al. (2004) however, shows a significant increase in plasma estradiol for japanese medaka females at 0.1 and 0.5 μg/L Fluoxetine exposure for four weeks.

In Sebire et al. (2015) male stickleback fish exposed to 32 μg/L of Fluoxetine showed a delayed response to a stimulus representing a threat and a decrease in aggressive behaviours. Similarly, the reproductive behaviour of male fathead minnows is significantly decreased when exposed to 0.1 μg/L Fluoxetine (Mennigen et al., 2011). The same species exposed to Fluoxetine at 1 μg/L, an environmentally relevant concentration found in freshwater systems, significantly affected mating behaviour, specifically nest building and defending in male fish (Weinberger and Klaper, 2014). This suggests that, similar effects in reproductive behaviours can occur at different concentrations and hence why, in accordance with the collected sample, there was no significant moderating effect of concentration on this class. These effects are also apparent in the overall Hedges' g estimate for Reproductive Behaviour ($g = -0.065$) and as Prichard and Granek (2016) note, changes in these behaviours may have implications for predator-prey interactions, food web dynamics, community composition/structure, and potentially biodiversity. However, Holmberg et al. (2011) for example, exposed rainbow trouts and guppies from environmentally relevant concentrations to high concentrations (1, 10, 100 μg/L) of citalopram (for 3 to 7 days) and did not detect significant effects on their sexual behaviour.

Some articles report results on Sexual Maturation that are also in conformity with the this study's overall effect (which was nonetheless non significant). Namely, Fluoxetine, Sertraline, Venlafaxine and Bupropion

at high concentrations caused a significant reduction in male fathead minnows' secundary sex characteristics (Schultz et al., 2011). One possible explanation is that exposure to antidepressants causes impaired testosterone production which is responsible for inducing the development of these characteristics (Schultz et al., 2011). Mosquitofish chronically exposed from juvenile stage to adult life at 71 μg/L showed a delayed development of external sexual morphology (gonopodium and black spot) showing that Fluoxetine affects sexual maturation of fish though at concentrations higher than the ones generally found in the environment (Henry and Black, 2008). Taking into account the positive overall effect on sexual maturation obtained ($g = 0.033$), one interesting endpoint measured, in Campos (2012), is the age at first reproduction, which according to the overall effect apparently increases with antidepressant exposure. This might be an indirect consequence, as the exposed organisms will have a larger investment in the reproductive structures (gonads, secondary sexual characteristics) which, in turn, only become fully matured at a later age. Thus, aquatic animals may have a delay in their first reproduction when compared to the unexposed organisms.

Some of these patterns are visible in the forest plots (Appendix C). In Figure 4.3, for instance, is represented the forest plot for Fecundity data (fish and crustacea) with Concentration taken into account as a moderator. Varano et al. (2017) tests the effects of Fluoxetine exposure on *Daphnia magna* fecundity and in the forest plot there is a clear increase in the negative effect of exposure on fecundity when the concentration of Venlafaxine increases, in both trials, which supports the negative estimate coefficient obtained for crustacea fecundity data in Table 3.6 ($g = -0.453$).

If we look at Figure C.2 we can see that the last study (Lister et al., 2009) points out an increase in the negative effect of exposure when the concentration increases. This result is contrary to what the moderator analysis coefficient estimate suggests for fish ($g = 0.368$). Being positive, the response should increase and not decrease with a rise in concentration. However, this can be explained by the fact that this study does not have as much weight as the others. Additionally, it has been argued that studies yielding relatively many effect sizes may excessively contribute to the meta-analytic results, as such, it should be noted that there are only 3 studies in this forest plot with the first two contributing with many more effect sizes than Lister et al. (2009).

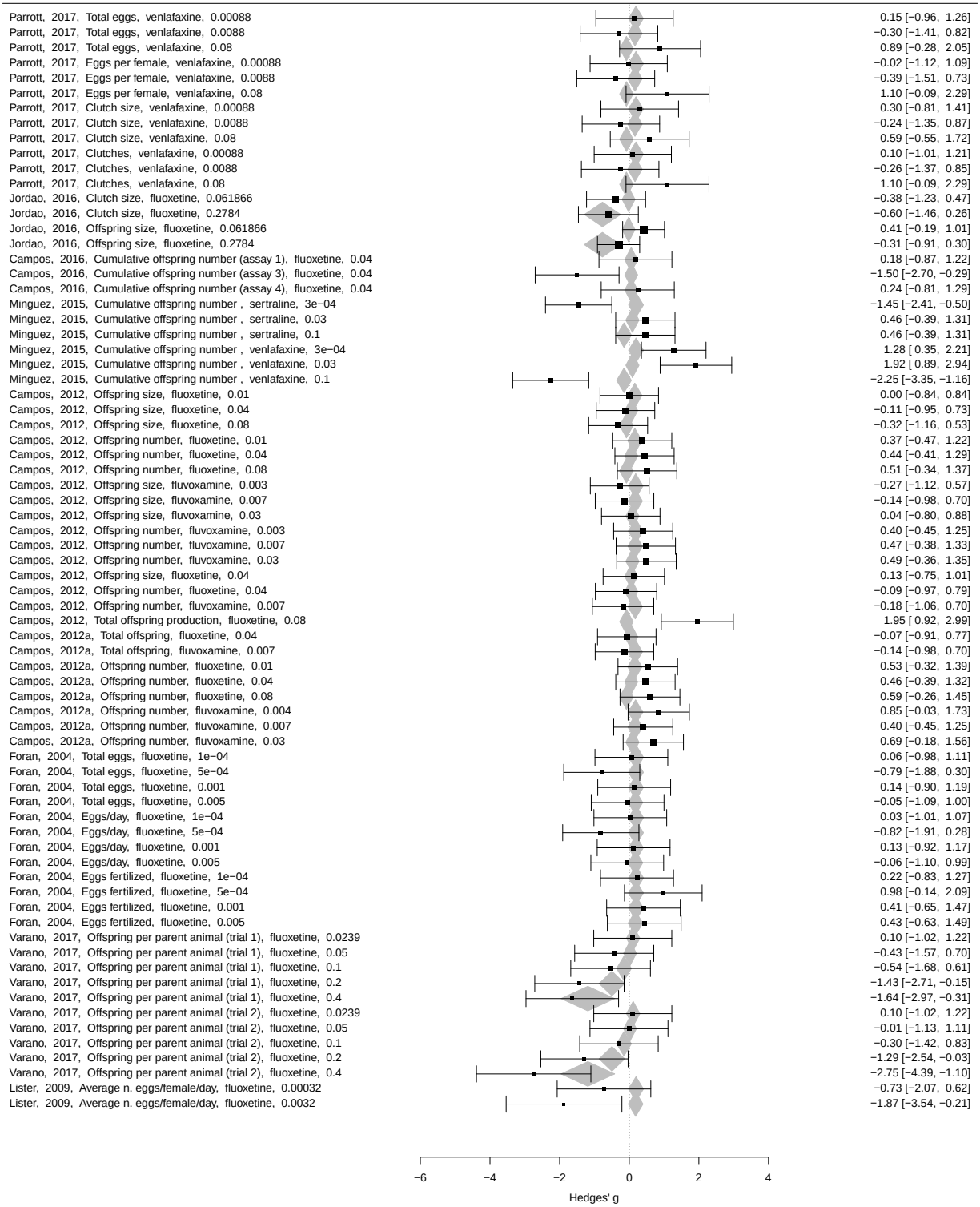


Figure 4.3: Fecundity forest plot with Concentration as moderator

4.3 Strengths, Limitations & Future Research

The present meta-analysis provides insight on the effects of antidepressants exposure on the reproduction of fish and crustaceans. However, only 19 articles were detected from the year 2000 to 2019, when selection criteria for this study were applied in the search process. Fish animal models used in the included published studies are all freshwater species, as well as *Daphnia magna* in the crustacean studies. Melvin and Wilson (2013), conducted a meta-analysis to compare studies using behavioural endpoints to others assessing acute lethality, development and reproductive endpoints, with results showing that the average duration of behavioural studies is consistently lower and that behavioural endpoints are generally more sensitive than reproductive and developmental endpoints. As such, behavioural studies should be warranted more attention has they are fast and sensitive tools for assessing toxicological effects of environmental contaminants, with particular interest in toxicants targeting the central nervous system.

Additionally, current toxicity data only encompasses short-term exposures (Santos et al., 2009), but aquatic organisms might be chronically exposed to various pharmaceuticals due to their continuous discharges, which, as time goes by, could bioaccumulate in their bodies, cause effects and change food-web properties. Therefore, despite coming at higher temporal and financial costs, there is a need for investigation regarding chronic exposure effects (Brodin et al., 2014; Corcoran et al., 2010; Kümmerer, 2010), namely, full life-cycle or multi-generational experiments that assess changes in fecundity, sexual maturation and reproductive behaviour of fish (Overturf et al., 2015), since changes in these endpoints do not become evident as fast as changes in molecular responses, for example.

As a result from the small number of studies, effect sizes and a lack of information on sample characteristics, moderator analysis was limited in the sense that there was still significant unexplained heterogeneity between effect sizes for the majority of subsets of the data which could have been explained by other moderator variables, such as temperature, pH or salinity, among others. Still regarding the size of the data, it has been reported that a small number of studies might result in underestimated standard errors and in biased estimates of the between-study variance when REstricted Maximum Likelihood (REML) method is used to estimate the parameters in the three-level meta analytic model (Van den Noortgate et al., 2013). In spite of this, since there are not yet methods available for determining the exact power in multilevel meta-analysis (Assink and Wibbelink, 2016), it was not possible to evaluate if the present meta-analysis was underpowered. Regardless, researchers should be aware of this limitation when interpreting the results. It is important to emphasize that these data limitations are similar for classic/narrative reviews, but the lack of quantitative approach may lead to these reviews' overestimation of effects from individual based results only.

This study, not only separates the measured reproductive endpoints into different classes, and analyses them separately, but also applies a three-level structure to the variance of the model. This avoids interdependency of effect sizes from the same study, so that all information is preserved and maximum statistical power can be achieved.

It is a rather recent method, which has not been widely applied in meta-analysis research. Following an example in Van den Noortgate et al. (2013), since this approach accounts for the hierarchical structure in the data, if one study results in 20 effect sizes for the analysis, this study will not contribute 20 times as much to

the estimation of the overall effect size when compared to another study which only contributes with 1 effect size. Rather, this "larger" study will yield information about one study-specific mean effect in the distribution of study mean effects. The weight attributed to each study will depend on the existing dependence between the effect sizes from the same study, i.e, the larger this dependence, the more the weight attributed to each effect size depends on the number of effect sizes reported in the study (Van den Noortgate et al., 2013).

Publication bias analysis were also conducted. The results showed that there was not a clear-cut indication of publication bias in any of the sets of data considered. Regarding funnel plots, it is debatable if they are useful when multilevel structures of underlying data are considered. For example, sets of points might cluster together because of statistical dependencies, which can be accounted for by using an appropriate multilevel model (Van den Noortgate et al., 2013), as the three-level meta-analytic model employed in this study. However, in a funnel plot (contour-enhanced or not), the inherent dependencies are not visible because the sets of points are just being displayed. They are only presented in this study to see if there is visual evidence of asymmetry in the association between sample size and effect size.

Usually, if Egger's test result is significant one would apply the Duval & Tweedie's trim-and-fill procedure (Duval and Tweedie, 2000), which estimates the overall effect if the missing "smaller" studies had been published. When it comes to assessment and handling of missing data, the available methods have not yet been evaluated in multi-level meta-analytic research, which makes it difficult to select a method for detecting and handling missing data in this project. The trim-and-fill method is an example of this, as it is yet to be extended to three-level meta-analytic models, which demonstrates the lack of tools available in multilevel meta-analytic research.

Despite this, trim-and-fill could be applied if a simple random effects model is considered. With this, one would obtain a funnel plot with the supposedly missing studies — in this case effect sizes — until symmetry is reached. Through this procedure it was evident that for funnel plots that showed significant asymmetry (Fecundity data of fish and crustaceans together, and Fecundity data regarding just crustaceans) there were effect sizes missing mostly in high significance areas, which might be indicative of factors other than publication bias, such as variable study quality. If the contour-enhanced funnel plot had not been considered, publication bias would have been assumed (Peters et al., 2008).

Another limitation to this analysis relates to when there are many effect sizes in the analysis, as it can be hard to represent them graphically. The forest plots for this study were obtained using the `forest` function of the `metafor` package, which does not have many arguments that allow the manipulation of the aspect to be more visually appealing, hence the lower quality of the forest plots here presented.

It is also important to note that, meta-analytic results should be interpreted in terms of biological importance and not statistical significance. This is the idea behind the term "effective thinking" (Nakagawa and Cuthill, 2007), as interpreting the overall effects only in terms of statistical significance might be wrong, since not rejecting the null-hypothesis does not necessarily mean that there is no real effect, but rather that there was not enough evidence to prove otherwise (Cohen, 1990). As such, the magnitudes of effect and their uncertainties should be placed above p-values and statistical significance (Nakagawa et al., 2017).

Finally, it is important to emphasize that researchers should aim to publish any non-significant results. Since

statistical power increases with the increase of the number of studies and effect sizes (Assink and Wibbelink, 2016), it is important to report every result even if unsignificant, as it is unfortunate not to have them in a meta-analysis.

4.4 Conclusion

Despite the role that pharmaceuticals in general have in human and animal lives today, little is known about their effects in aquatic wildlife once they are released into the environment. This study aimed to provide more insight into the association between pharmaceutical exposure, specifically antidepressants, and it's effects on the reproduction of aquatic organisms. While 15 out of the 19 included studies reported significant results individually, the findings of this multi-level meta-analytic approach suggest that, overall, there is no significant relation between antidepressant exposure and reproductive measures of fish and crustaceans. Usually, narrative reviews report significant effects (in general), because the individual studies reported significant effects. This work shows the same is not verified when we compare studies via a quantitative analysis (with standardized responses), as it does not support the idea that there are general effects in reproduction. However, when concentration is taken into account there is a significant association between antidepressant exposure and fecundity in fish and crustaceans. These two groups have, curiously, different outcomes with concentration having a positive effect on fish fecundity and a negative effect on crustacean fecundity. This work also evidenced data limitations when considering, for example, toxicants and species; with the majority of studies focusing on one compound (Fluoxetine) and only freshwater species. As such, this meta-analytic review contributes to the literature on ecotoxicological effects of pharmaceutical exposure in aquatic organisms, providing a quantitative approach applicable in other ecotoxicological studies.

Bibliography

Aminot, Y., Le Menach, K., Pardon, P., Etcheber, H., and Budzinski, H. (2016). Inputs and seasonal removal of pharmaceuticals in the estuarine Garonne River. *Marine Chemistry*, 185:3–11.

Arnold, K. E., Brown, A. R., Brown, A. R., Ankley, G. T., and Sumpter, J. P. (2014). Medicating the environment: Assessing risks of pharmaceuticals to wildlife and ecosystems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1656).

Assink, M. and Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. 12(3):154–174.

aus der Beek, T., Weber, F. A., Bergmann, A., Hickmann, S., Ebert, I., Hein, A., and Küster, A. (2016). Pharmaceuticals in the environment-Global occurrences and perspectives. *Environmental Toxicology and Chemistry*, 35(4):823–835.

Bakkalbasi, N., Bauer, K., Glover, J., and Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Digital Libraries*.

Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., Wallet, P. A., Fiset, M., and Huang, B. (2004). How Does Distance Education Compare to Classroom Instruction? *Review of Educational Research*.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2009). *Introduction to Meta-Analysis*.

Brodin, T., Klaminder, J., Piovano, S., Jonsson, M., Fick, J., and Heynen, M. (2014). Ecological effects of pharmaceuticals in aquatic systems–impacts through behavioural alterations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1656):20130580–20130580.

Campos, B. (2012). Mechanisms of Action of Selective Serotonin Reuptake Inhibitors in Daphnia magna.

Campos, B., Rivetti, C., Kress, T., Barata, C., and Dircksen, H. (2016). Depressing Antidepressant: Fluoxetine Affects Serotonin Neurons Causing Adverse Reproductive Responses in Daphnia magna. *Environmental Science and Technology*, 50(11):6000–6007.

Chen, D. G. D. and Peace, K. E. (2013). *Applied Meta-Analysis with R*.

Cheung, M. W. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, 19(2):211–229.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates, 2nd edition.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*.

Corcoran, J., Winter, M. J., and Tyler, C. R. (2010). Pharmaceuticals in the aquatic environment: A critical review of the evidence for health effects in fish. *Critical Reviews in Toxicology*, 40(4):287–304.

Crane, M., Watts, C., and Boucard, T. (2006). Chronic aquatic environmental risks from exposure to human pharmaceuticals.

Crombie, I. K. and Davies, H. T. O. (2009). What is a meta-analisys? 2nd(April):1–8.

Daughton, C. G. (2014). The Matthew Effect and widely prescribed pharmaceuticals lacking environmental monitoring: Case study of an exposure-assessment vulnerability. *Science of the Total Environment*.

Daughton, C. G. and Ternes, T. A. (1999). Pharmaceuticals and personal care products in the environment: Agents of subtle change?

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3):177–188.

Donnachie, R. L., Johnson, A. C., and Sumpter, J. P. (2016). A rational approach to selecting and ranking some pharmaceuticals of concern for the aquatic environment and their relative importance compared with other chemicals. *Environmental Toxicology and Chemistry*.

Du, B., Haddad, S. P., Luek, A., Scott, W. C., Saari, G. N., Burket, S. R., Breed, C. S., Kelly, M., Broach, L., Rasmussen, J. B., Chambliss, C. K., and Brooks, B. W. (2016). Bioaccumulation of human pharmaceuticals in fish across habitats of a tidally influenced urban bayou. *Environmental Toxicology and Chemistry*.

Duval, S. and Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*.

Edens, J. F., Campbell, J. S., and Weir, J. M. (2007). Youth psychopathy and criminal recidivism: A meta-analysis of the psychopathy checklist measures. *Law and Human Behavior*.

Egger, M., Smith, G. D., Schneider, M., and Minder, C. (1997). Bias in Meta-analysis Detected By a Simple , Graphical Test. 14:1–12.

European Commission (2013). Directive 2013/39/EU of the european parliament and of the council of 12 August 2013. *Official Journal of the European Union*, 24.8.2013(July):L 226/1 – L 226/17.

European Parliament (2008). Directive 2008/105/EC.

Falchikov, N. and Goldfinch, J. (2000). Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks. *Review of Educational Research*.

Fatta-Kassinos, D., Meric, S., and Nikolaou, A. (2011). Pharmaceutical residues in environmental waters and wastewater: Current state of knowledge and future research.

Fent, K., Weston, A. A., and Caminada, D. (2006). Ecotoxicology of human pharmaceuticals. *Aquatic Toxicology*, 76(2):122–159.

Flaherty, C. M. and Dodson, S. I. (2005). Effects of pharmaceuticals on Daphnia survival, growth, and reproduction. *Chemosphere*.

Fong, P. P. and Ford, A. T. (2014). The biological effects of antidepressants on the molluscs and crustaceans: A review. *Aquatic Toxicology*, 151:4–13.

Fonseca, V. and Reis-Santos, P. (2018). 1,2 1 , 2. pages 473–481.

Foran, C. M., Weston, J., Slattery, M., Brooks, B. W., and Huggett, D. B. (2004). Reproductive Assessment of Japanese Medaka (Oryzias latipes) Following a Four-Week Fluoxetine (SSRI) Exposure. *Archives of Environmental Contamination and Toxicology*, 46(4):511–517.

Furuhagen, S., Fuchs, A., Belleza, E. L., Breitholtz, M., and Gorokhova, E. (2014). Are pharmaceuticals with evolutionary conserved molecular drug targets more potent to cause toxic effects in non-target organisms? *PLoS ONE*.

Gaw, S., Thomas, K. V., and Hutchinson, T. H. (2014). Sources, impacts and trends of pharmaceuticals in the marine and coastal environment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1656).

Glassmeyer, S. T., Kolpin, D. W., Furlong, E. T., and Focazio, M. J. (2007). Environmental presence and persistence of pharmaceuticals an overview. In *Fate of Pharmaceuticals in the Environment and in Water Treatment Systems*.

Harrer, M. and Ebert, D. (2018). Doing Meta-Analysis in R: A practical Guide. https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/. Accessed in 18-04-19.

Hausknecht, J. P., Day, D. V., and Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*.

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., and Jennions, M. D. (2015). The Extent and Consequences of P-Hacking in Science. *PLoS Biology*.

Hedges, L. V. (2008). Distribution Theory for Glass's Estimator of Effect size and Related Estimators. *Journal of Educational Statistics*.

Henry, T. B. and Black, M. C. (2008). Acute and chronic toxicity of fluoxetine (selective serotonin reuptake inhibitor) in western Mosquitofish. *Archives of Environmental Contamination and Toxicology*.

Higgins, J. P. T. (2003). Measuring inconsistency in meta-analyses. *BMJ*, pages 557–560.

Higgins, S., Wilson, J. Y., Purdy, J., Li, H., Chow, J., Kirischian, N., Galus, M., Rangaranjan, S., and Metcalfe, C. (2013). Chronic, low concentration exposure to pharmaceuticals impacts multiple organ systems in zebrafish. *Aquatic Toxicology*, 132-133:200–211.

Holmberg, A., Fogel, J., Albertsson, E., Fick, J., Brown, J. N., Paxéus, N., Förlin, L., Johnsson, J. I., and Larsson, D. G. (2011). Does waterborne citalopram affect the aggressive and sexual behaviour of rainbow trout and guppy? *Journal of Hazardous Materials*, 187(1-3):596–599.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications: Second edition*.

Huedo-Medina, T. B., Sanchez-Meca, F., Marin-Martinez, F., and Botella, J. (2006). Assessing heterogeneity in meta-analysis: I2 or Q statistic? *Psychological Methods*, 11:193–206.

Huerta, B., Rodríguez-Mozaz, S., and Barceló, D. (2012). Pharmaceuticals in biota in the aquatic environment: Analytical methods and environmental implications. *Analytical and Bioanalytical Chemistry*, 404(9):2611–2624.

Hunter, J. E., Schmidt, F. L., and Raudenbush, S. W. (1991). Methods of Meta-Analysis: Correcting Error and Bias in Research Findings. *Journal of the American Statistical Association*.

Knapp, G. and Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, 22(17):2693–2710.

Kümmerer, K. (2003). Promoting resistance by the emission of antibiotics from hospitals and households into effluent. *Clinical Microbiology and Infection*, 9(12):1203–1214.

Kümmerer, K. (2010). Pharmaceuticals in the environment. *Annual Review of Environment and Resources*, 63(2):174–178.

Länge, R., Hutchinson, T. H., Croudace, C. P., Siegmund, F., Schweinfurth, H., Hampe, P., Panter, G. H., and Sumpter, J. P. (2001). Effects of the synthetic estrogen 17α-ethinylestradiol on the life-cycle of the fathead minnow (Pimephales promelas). *Environmental Toxicology and Chemistry*.

Larsson, D. G., de Pedro, C., and Paxeus, N. (2007). Effluent from drug manufactures contains extremely high levels of pharmaceuticals. *Journal of Hazardous Materials*.

Lièvre, M., Cucherat, M., and Leizorovicz, A. (2002). Pooling, meta-analysis, and the evaluation of drug safety. *Current Controlled Trials in Cardiovascular Medicine*.

Lister, A., Regan, C., Van Zwol, J., and Van Der Kraak, G. (2009). Inhibition of egg production in zebrafish by fluoxetine and municipal effluents: A mechanistic evaluation. *Aquatic Toxicology*, 95(4):320–329.

Long, J. (2001). An Introduction to and Generalization of the "Fail-Safe N.". *Paper presented at the Annual Meeting of the Southwest Educational Research Association (New Orleans, LA, February1-3, 2001)*.

Maskaoui, K. and Zhou, J. L. (2010). Colloids as a sink for certain pharmaceuticals in the aquatic environment. *Environmental Science and Pollution Research*.

Melvin, S. D. and Wilson, S. P. (2013). Chemosphere The utility of behavioral studies for aquatic toxicology testing :. *Chemosphere*, 93(10):2217–2223.

Mennigen, J. A., Lado, W. E., Zamoraa, J. M., Duarte-Guterman, P., Langlois, V. S., Metcalfe, C. D., Chang, J. P., Moon, T. W., and Trudeau, V. L. (2010). Waterborne fluoxetine disrupts the reproductive axis in sexually mature male goldfish, Carassius auratus. *Aquatic Toxicology*, 100(4):354–364.

Mennigen, J. A., Stroud, P., Zamora, J. M., Moon, T. W., and Trudeau, V. L. (2011). Pharmaceuticals as neuroendocrine disruptors: Lessons learned from fish on prozac. *Journal of Toxicology and Environmental Health - Part B: Critical Reviews*, 14(5-7):387–412.

Mennigen, J. A., Zamora, J. M., Chang, J. P., and Trudeau, V. L. (2017). Endocrine disrupting effects of waterborne fluoxetine exposure on the reproductive axis of female goldfish, Carassius auratus. *Comparative Biochemistry and Physiology Part - C: Toxicology and Pharmacology*, 202:70–78.

Nakagawa, S. and Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews*, 82(4):591–605.

Nakagawa, S., Noble, D. W., Senior, A. M., and Lagisz, M. (2017). Meta-evaluation of meta-analysis: Ten appraisal questions for biologists. *BMC Biology*, 15(1):1–14.

Nałecz-Jawecki, G. (2007). Evaluation of the in vitro biotransformation of fluoxetine with HPLC, mass spectrometry and ecotoxicological tests. *Chemosphere*.

Nicholls, R. J., Wong, P. P., Burkett, V., Codignotto, J., Hay, J., McLean, R., Ragoonaden, S., and Woodroffe, C. (2007). Coastal systems and low-lying areas. Climate Change 2007: Impacts, Adaptation and Vulnerability. *IPCC Report*, pages 315–356.

Orwin, R. G. (2008). A Fail-Safe N for Effect Size in Meta-Analysis . *Journal of Educational Statistics*.

Overturf, M. D., Anderson, J. C., Pandelides, Z., Beyger, L., and Holdway, D. A. (2015). Pharmaceuticals and personal care products: A critical review of the impacts on fish reproduction. *Critical Reviews in Toxicology*, 45(6):492–530.

Parrott, J. L. and Metcalfe, C. D. (2017). Assessing the effects of the antidepressant venlafaxine to fathead minnows exposed to environmentally relevant concentrations over a full life cycle. *Environmental Pollution*, 229:403–411.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., and Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, 61(10):991–996.

Pigott, T. D. (2012). *Advances in meta-analysis*.

Prichard, E. and Granek, E. F. (2016). Effects of pharmaceuticals and personal care products on marine organisms: from single-species studies to an ecosystem-based approach. *Environmental Science and Pollution Research*, 23(22):22365–22384.

Reis-Santos, P., Pais, M., Duarte, B., Caçador, I., Rosa, J., Freitas, A., Caçador, I., Ramos, F., Gillanders, B. M., Cabral, H. N., Barbosa, J., Leston, S., Fonseca, V. F., and Vila Pouca, A. S. (2018). Screening of human and veterinary pharmaceuticals in estuarine waters: A baseline assessment for the Tejo estuary. *Marine Pollution Bulletin*, 135(March):1079–1084.

Rothstein, H. R., Sutton, A. J., and Borenstein, M. (2006). *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*.

RStudio Team (2015). RStudio: Integrated Development Environment for R. http://www.rstudio.com/%7D.

Santos, L. H., Fachini, A., Pena, A., Araújo, A., Delerue-Matos, C., and Montenegro, M. (2009). Ecotoxicological aspects related to the presence of pharmaceuticals in the aquatic environment. *Journal of Hazardous Materials*, 175(1-3):45–95.

Schultz, M. M., Painter, M. M., Bartell, S. E., Logue, A., Furlong, E. T., Werner, S. L., and Schoenfuss, H. L. (2011). Selective uptake and biological consequences of environmentally relevant antidepressant pharmaceutical exposures on male fathead minnows. *Aquatic Toxicology*, 104(1-2):38–47.

Schuster, A., Hädrich, C., and Kümmerer, K. (2008). Flows of active pharmaceutical ingredients originating from health care practices on a local, regional, and nationwide level in Germany-is hospital effluent treatment an effective approach for risk reduction? *Water, Air, and Soil Pollution: Focus*, 8(5-6):457–471.

Sebire, M., Davis, J. E., Hatfield, R., Winberg, S., and Katsiadakia, I. (2015). Prozac affects stickleback nest quality without altering androgen , spiggin or aggression levels during a 21-day breeding test Dear Author , added to the article before publication , but are not reflected. 168(SEPTEMBER):78–89.

Seto, K. C., Guneralp, B., and Hutyra, L. R. (2012). Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences*.

Simonsohn, U., Nelson, L. D., and Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*.

Sterne, J. A., Egger, M., and Smith, G. D. (2001). Investigating and dealing with publication and other biases in meta-analysis. *BMJ (Clinical research ed.)*, 323(7304):101–5.

Sterne, J. A. C. and Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, chapter 6, pages 99–100.

Sutton, A. J. and Higgins, J. P. (2008). Recent developments in meta-analysis.

Tabachnick, B. G. and Fidell, L. S. (2014). *Using multivariate statistics (6th ed.)*.

The European Commission (2015). COMMISSION IMPLEMENTING DECISION (EU) 2015/ 495. *Official Journal of the European Union*.

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., and Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, 45(2):576–594.

Varano, V., Fabbri, E., and Pasteris, A. (2017). Assessing the environmental hazard of individual and combined pharmaceuticals: acute and chronic toxicity of fluoxetine and propranolol in the crustacean Daphnia magna. *Ecotoxicology*, 26(6):711–728.

Vesterinen, H. M., Sena, E. S., Egan, K. J., Hirst, T. C., Churolov, L., Currie, G. L., Antonic, A., Howells, D. W., and Macleod, M. R. (2014). Meta-analysis of data from animal studies: A practical guide. *Journal of Neuroscience Methods*, 221:92–102.

Viechtbauer, W. (2007). Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. *Journal of Educational and Behavioral Statistics*, 30(3):261–293.

Viechtbauer, W. (2015). Conducting Meta-Analyses in R with the metafor Package . *Journal of Statistical Software*.

Weinberger, J. and Klaper, R. (2014). Environmental concentrations of the selective serotonin reuptake inhibitor fluoxetine impact specific behaviors involved in reproduction, feeding and predator avoidance in the fish Pimephales promelas (fathead minnow). *Aquatic Toxicology*, 151:77–83.

Westhof, L., Köster, S., and Reich, M. (2016). Occurrence of micropollutants in the wastewater streams of cruise ships. *Emerging Contaminants*, 2(4):178–184.

Wilson, D. B. (2001). Meta-Analytic Methods for Criminology. *The Annals of the American Academy of Political and Social Science*.

Yunlong, L., Wenshan, G., Huu Hao, N., Long Duc, N., Faisal Ibney, H., Jian, Z., and Shuang, L. (2014). A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment. *Science of the Total Environment*, 473-474(March):619–641.

Ziegler, S., Koch, A., and Victor, N. (2001). Deficits and remedy of the standard random effects methods in meta-analysis. In *Methods of Information in Medicine*.

Appendices

Appendix A

Tables

Table A.1: Studies included in the analysis

Study	Organism Group	Endpoint Class	DOI
Campos, B. (2012a)	Crustacea	F; SM	10.1016/j.aquatox.2011.12.003
Campos, B. et al. (2012)	Crustacea	F; SM	10.1021/es203157f
Campos, B. et al. (2016)	Crustacea	F	10.1021/acs.est.6b00826
Dzieweczynski, TL. and Hebert, OL. (2012)	Fish	RB	10.1016/j.physbeh.2012.06.007
Foran, CM. et al. (2004)	Fish	F; H; MR; RB; SM	10.1007/s00244-003-3042-5
Henry, TB. and Black, MC. (2008)	Fish	SM	10.1007/s00244-007-9018-0
Higgins, S. et al. (2013)	Fish	ML	10.1016/j.aquatox.2012.12.021
Holmberg, A. et al. (2011)	Fish	RB	10.1016/j.jhazmat.2011.01.055
Jordão, R. et al. (2016)	Crustacea	F	10.1016/j.scitotenv.2015.12.097
Lister, A. et al. (2009)	Fish	F; MR	10.1016/j.aquatox.2009.04.011
Mennigen, J. et al. (2010)	Fish	MR	10.1016/j.aquatox.2010.08.016
Mennigen, J. et al. (2017)	Fish	MR; SM	10.1016/j.cbpc.2017.08.003
Minguez, L. et al. (2015)	Crustacea	F	10.1021/es504808g
Olsén, K. H. et al. (2014)	Fish	RB	10.1016/j.aquatox.2014.02.011
Parrott, J. L. and Metcalfe, C. D. (2017)	Fish	F; H; SM	10.1016/j.envpol.2017.06.011
Schultz, MM. et al. (2011)	Fish	MR; SM	10.1016/j.aquatox.2011.03.011
Sebire, M. et al. (2015)	Fish	MR; RB	10.1016/j.aquatox.2015.09.009
Varano, V. et al. (2017)	Crustacea	F	10.1007/s10646-017-1803-6
Weinberger, J. and Klaper, R. (2014)	Fish	RB	10.1016/j.aquatox.2013.10.012

Notes: F = Fecundity, H = Hatchability, ML = Molecular Responses, RB = Reproductive Behaviour, SM = Sexual Maturation.

Appendix B

Boxplots

Figure B.1: Concentration distribution boxplot

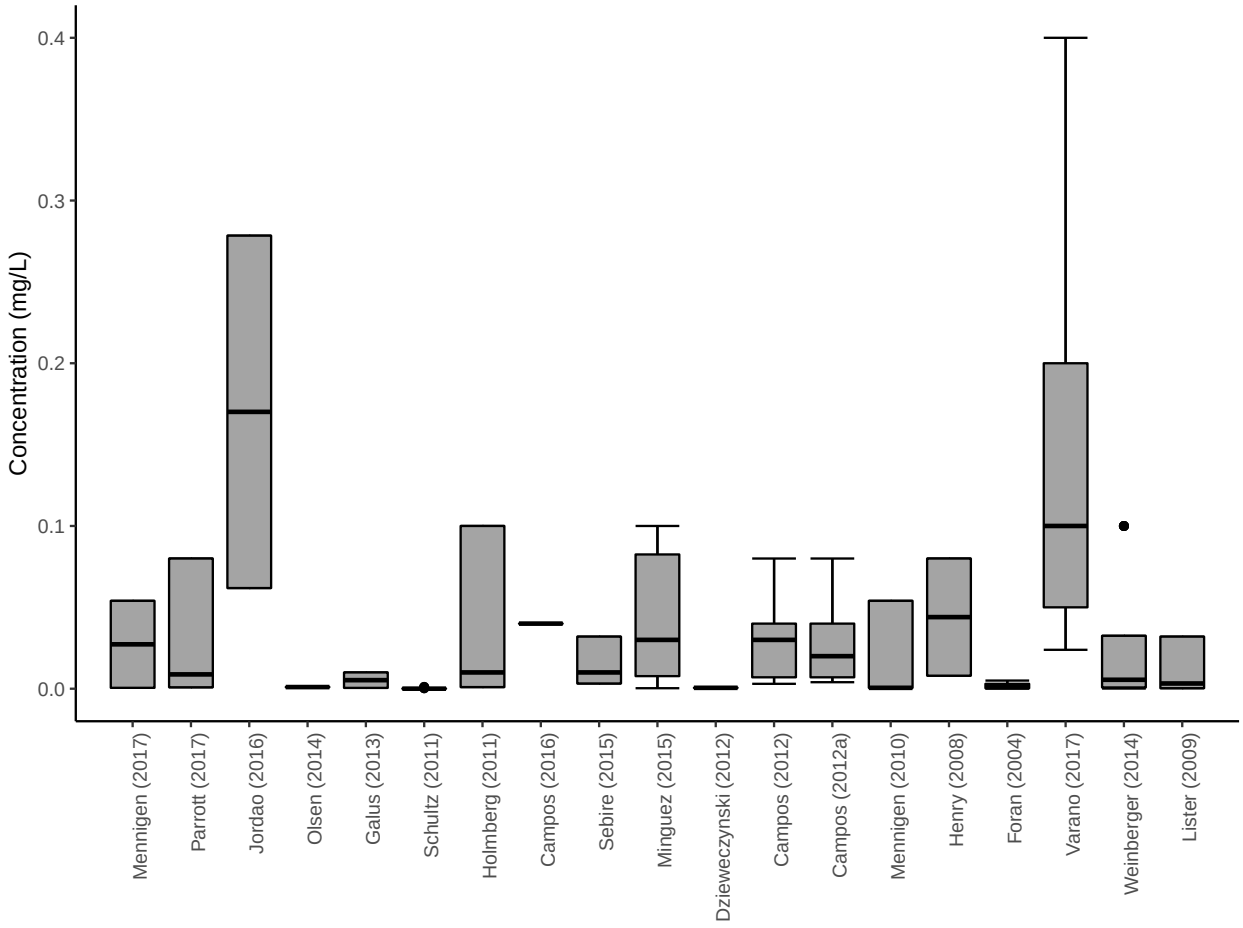
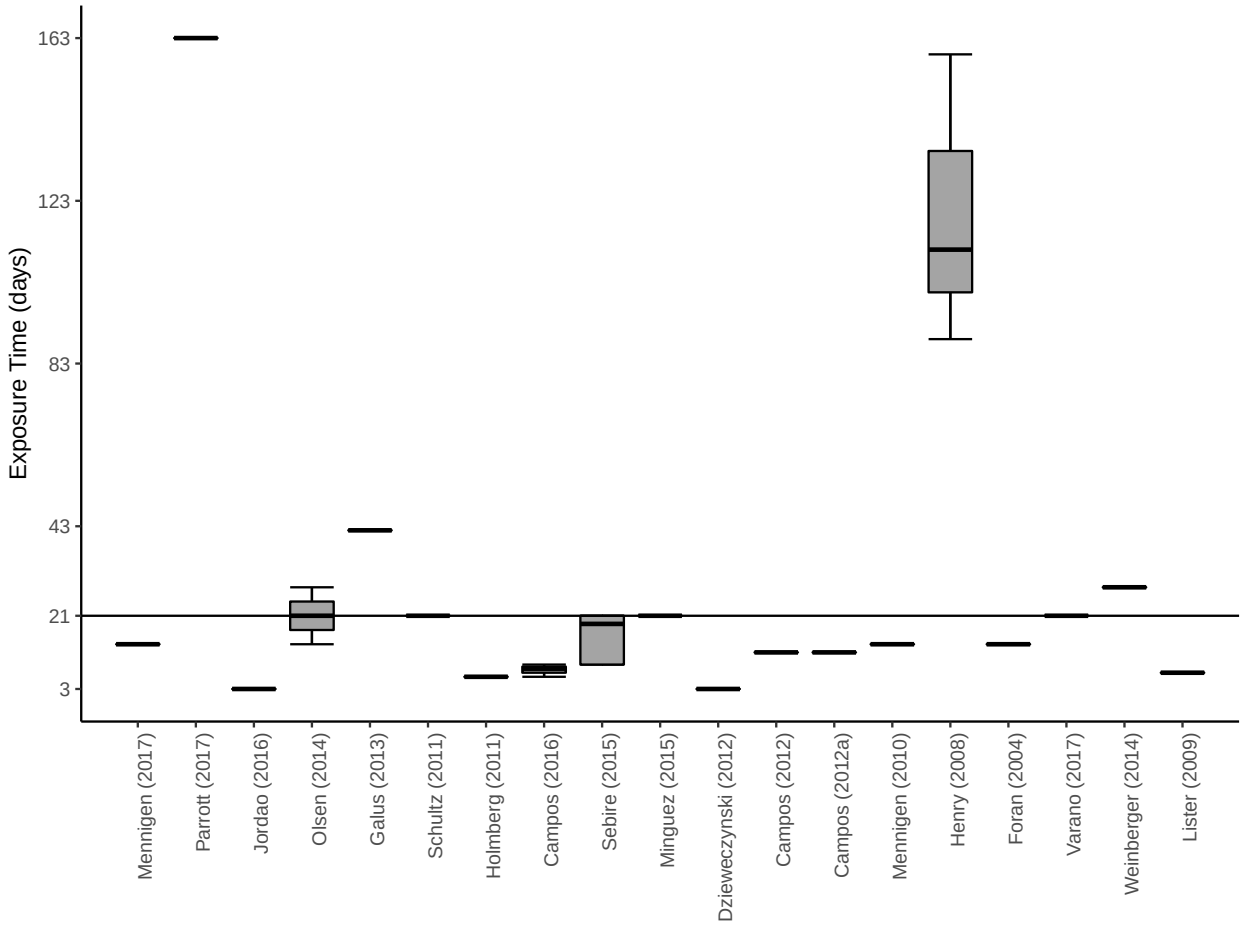


Figure B.2: Exposure Time distribution boxplot. Studies over the horizontal line crossing the y-axis at 21 days indicate that the organisms were chronically exposed.



Appendix C

Forest Plots

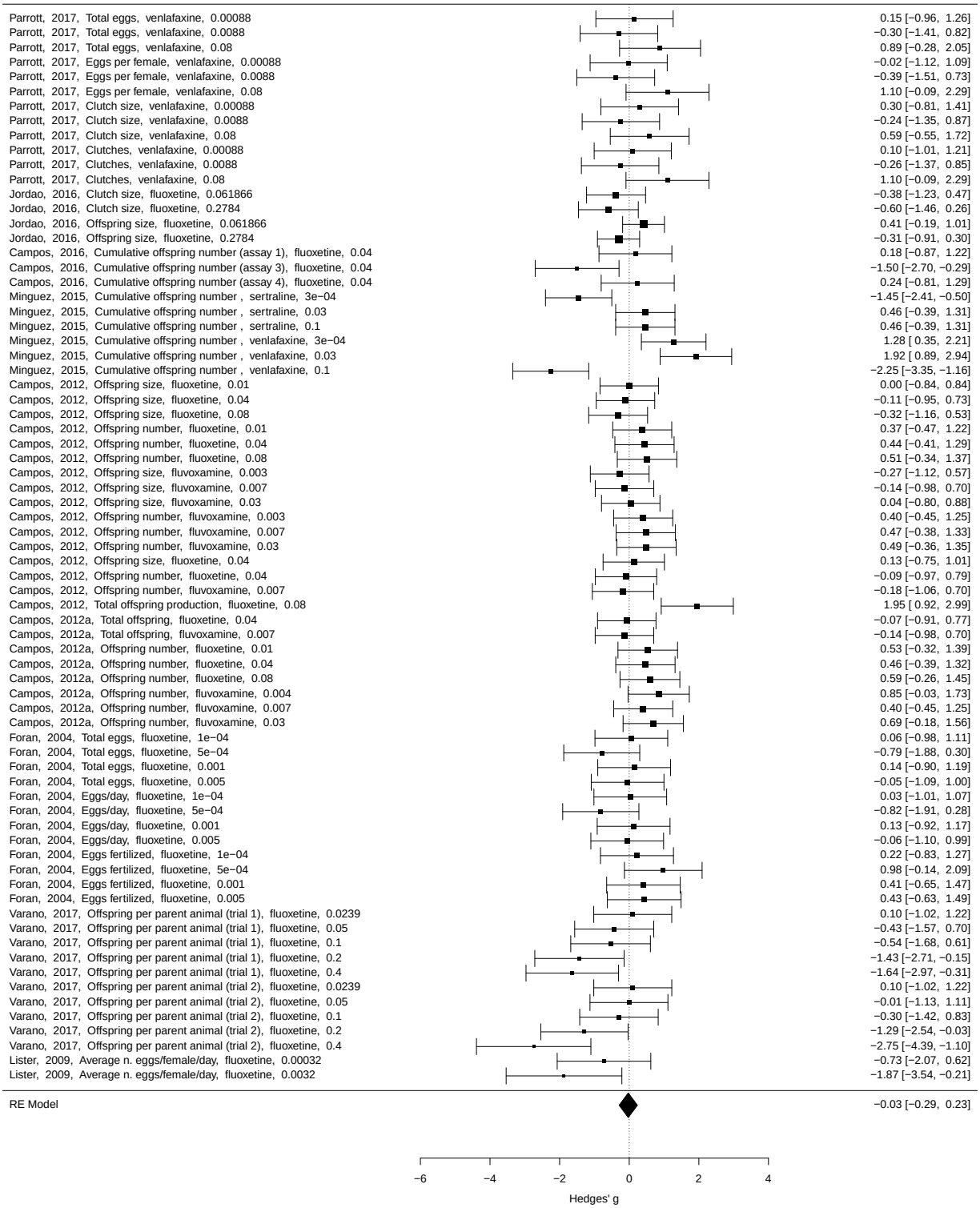


Figure C.1: Fecundity forest plot

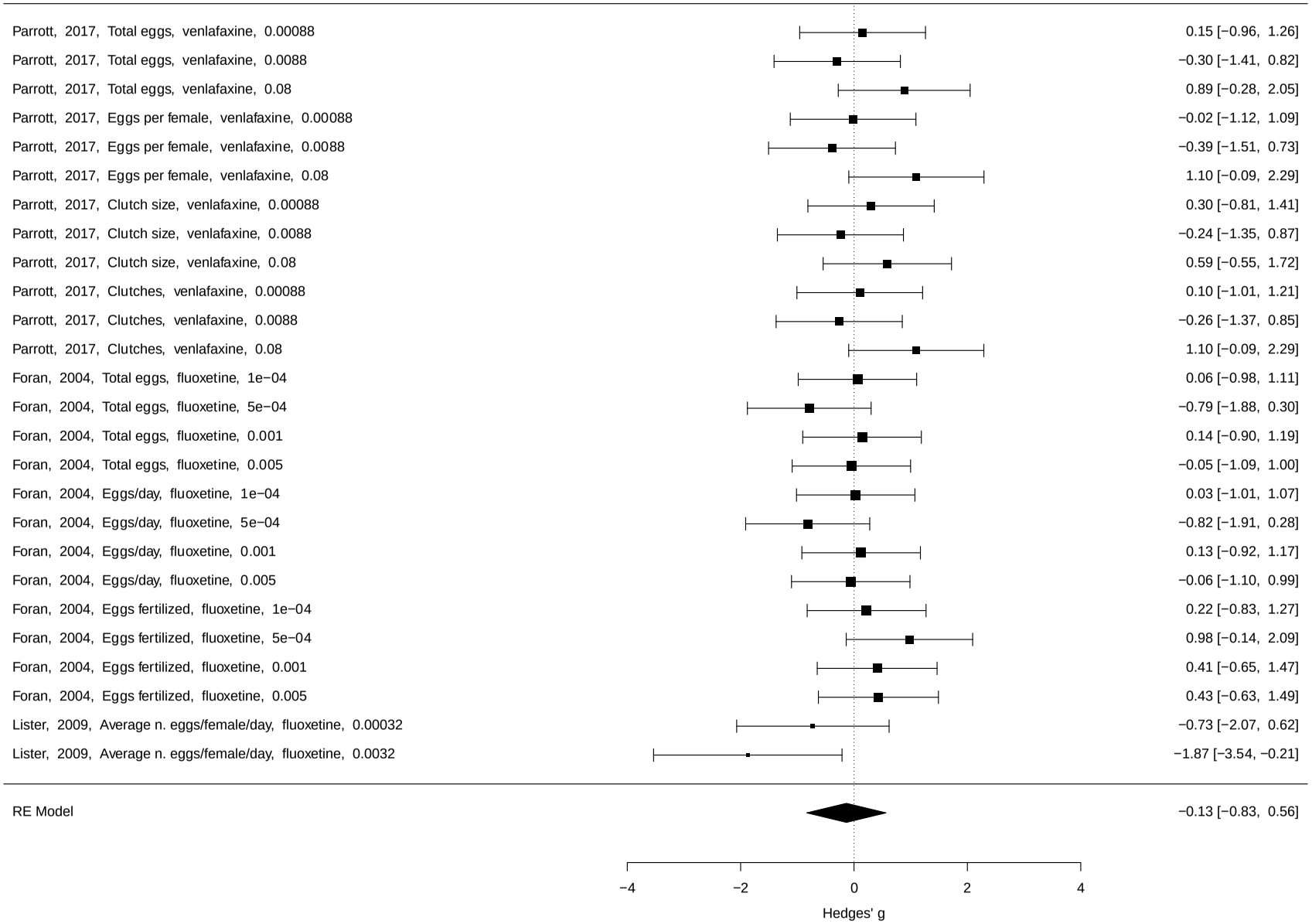


Figure C.2: Fecundity - fish forest plot

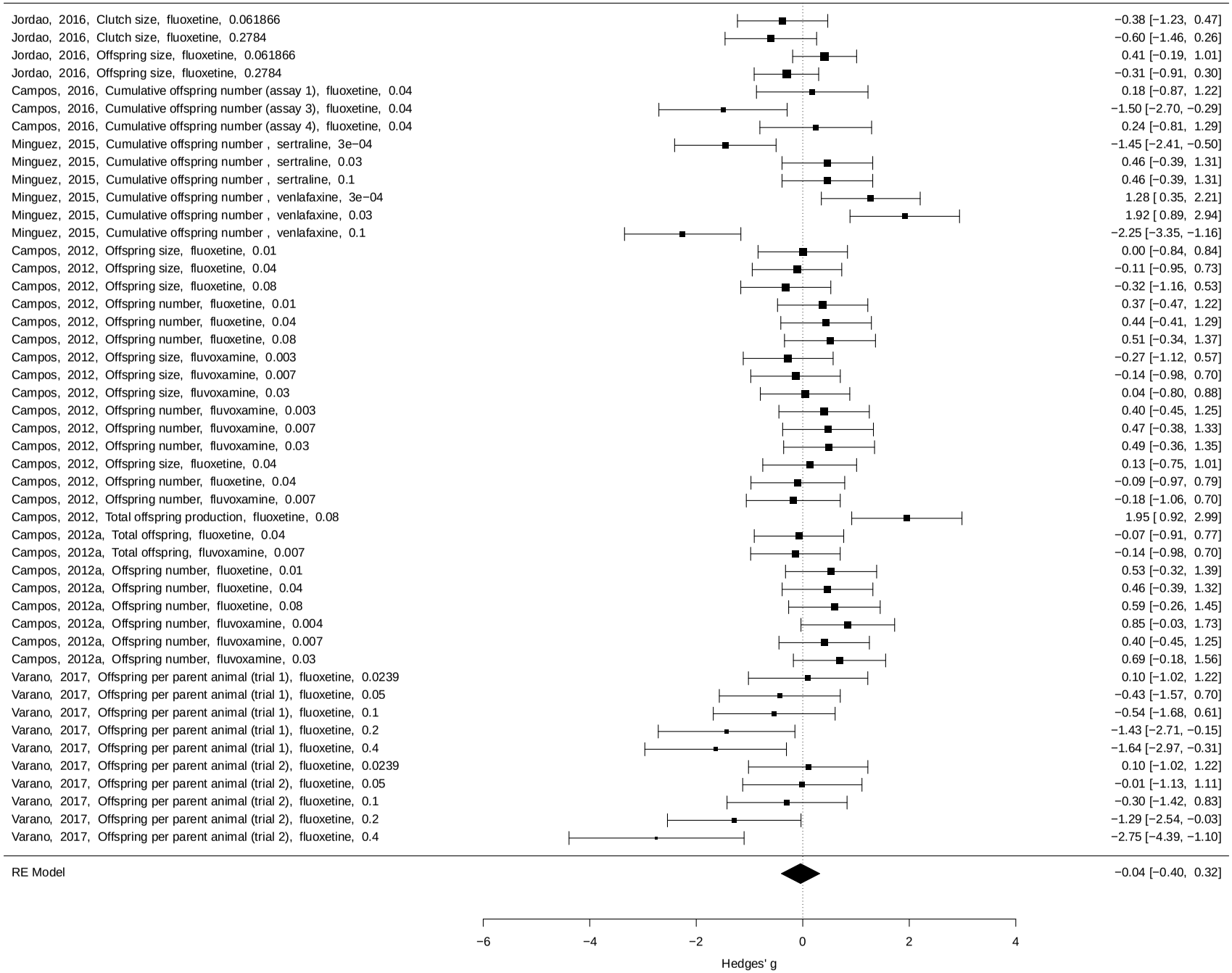


Figure C.3: Fecundity - crustacea forest plot

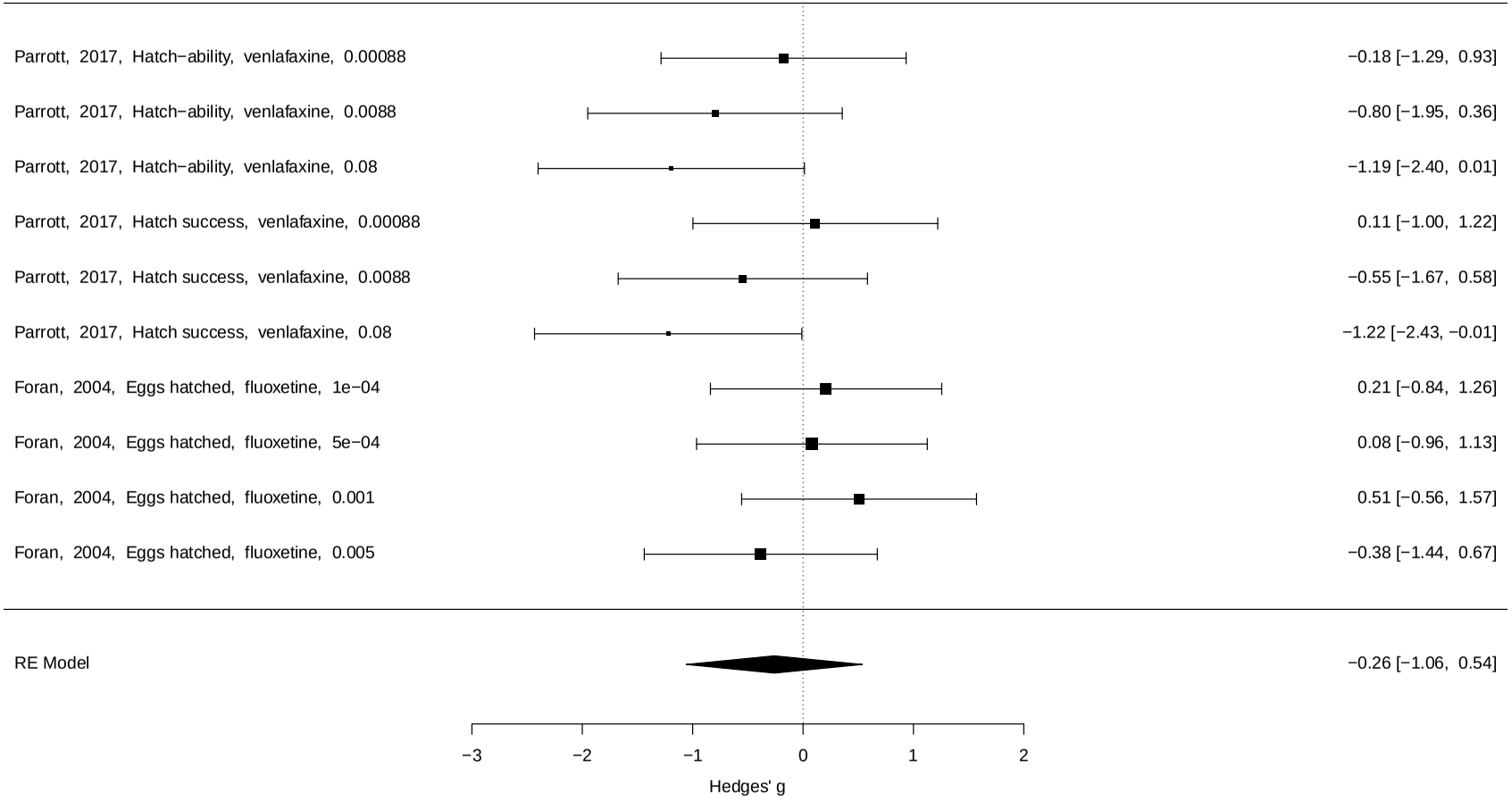


Figure C.4: Hatchability forest plot

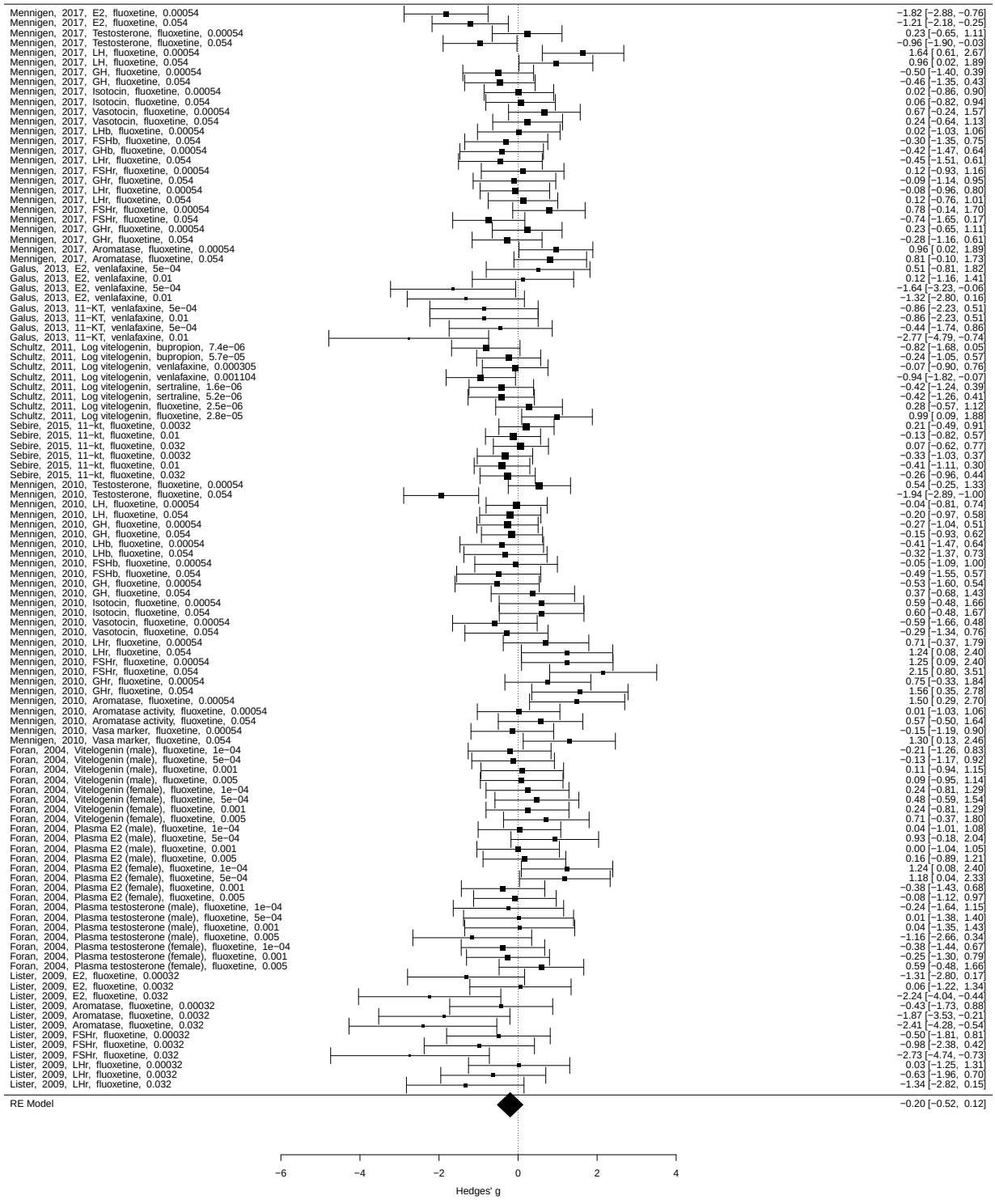


Figure C.5: Molecular Responses forest plot

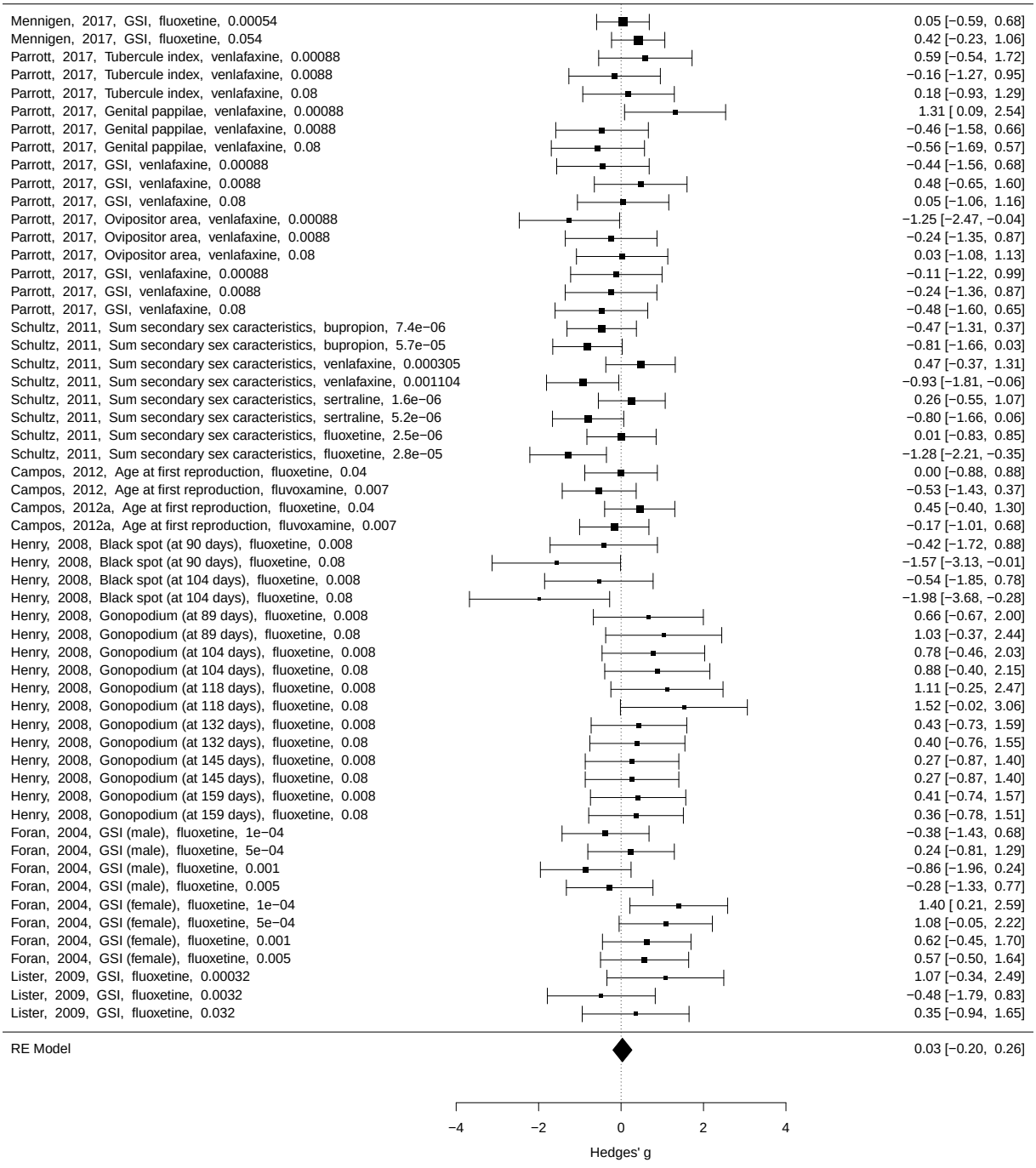


Figure C.6: Sexual Maturation forest plot

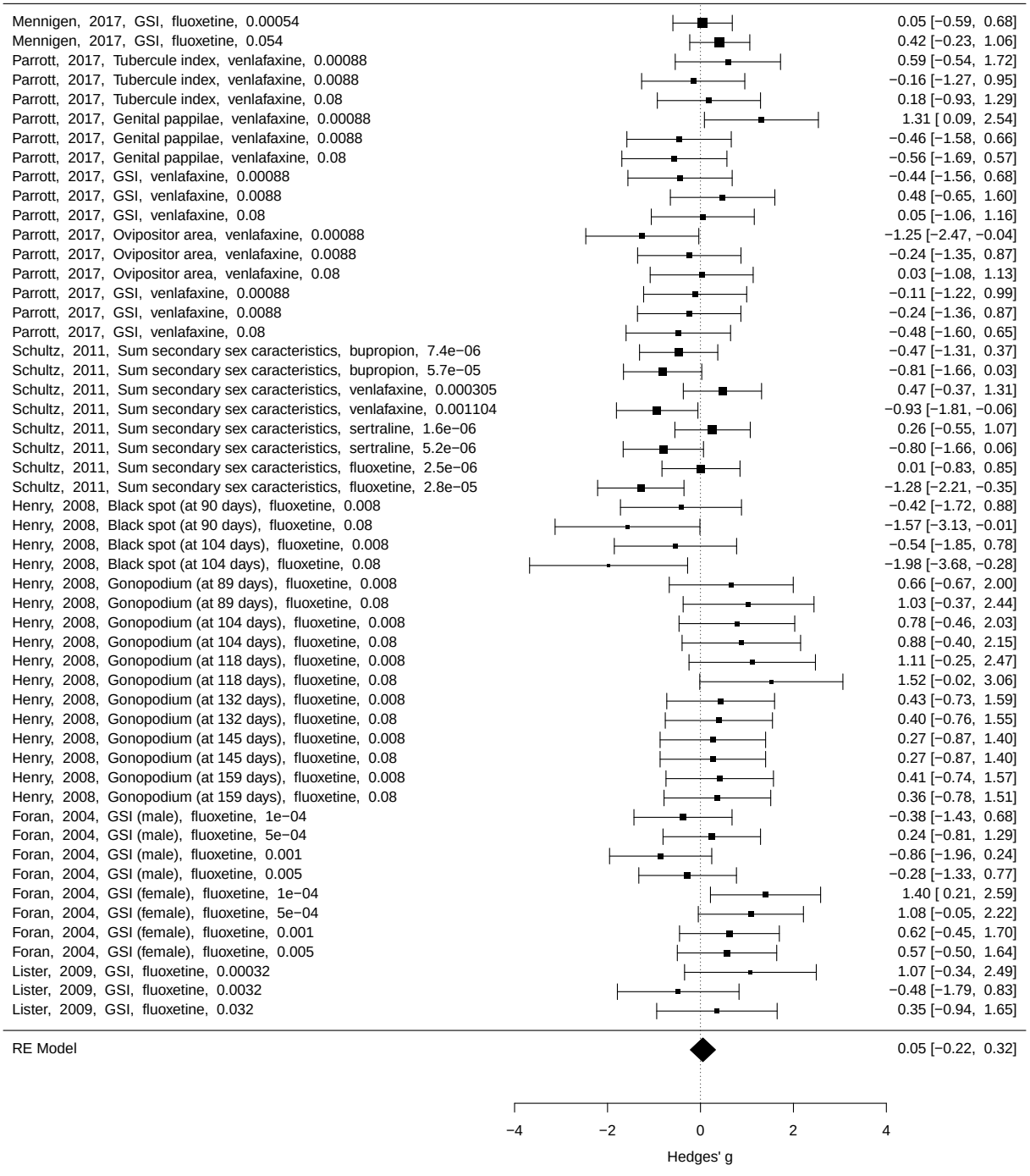


Figure C.7: Sexual Maturation - fish forest plot

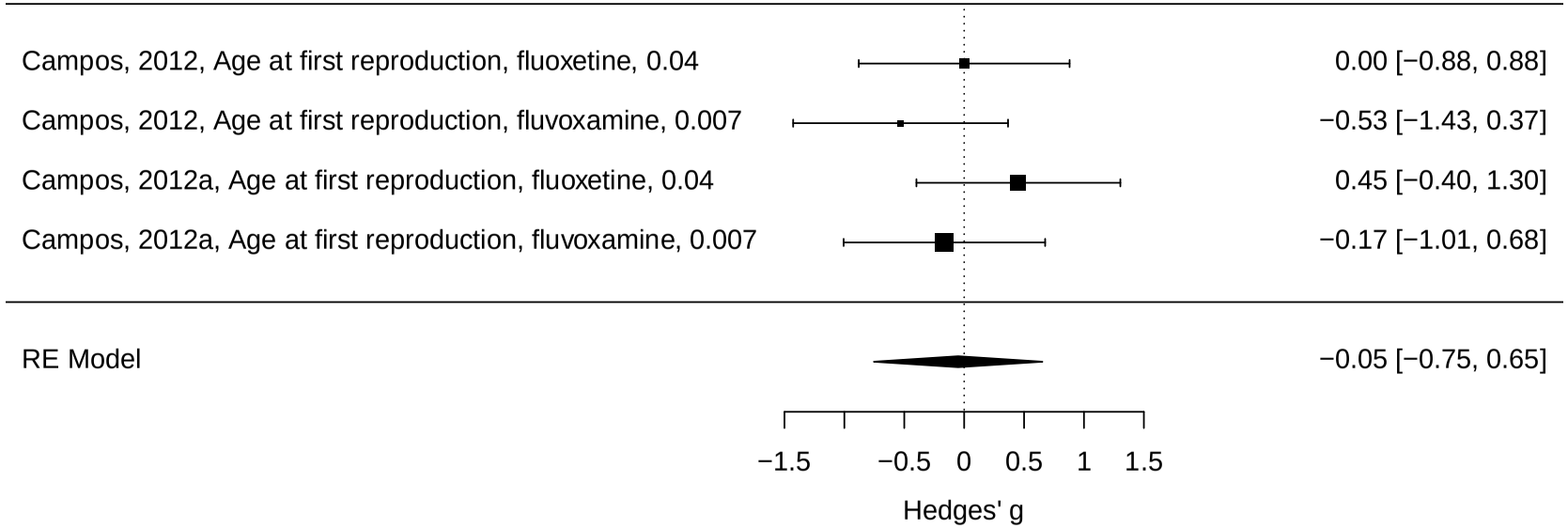


Figure C.8: Sexual Maturation - crustacea forest plot

Appendix D

Variance Distribution Plots

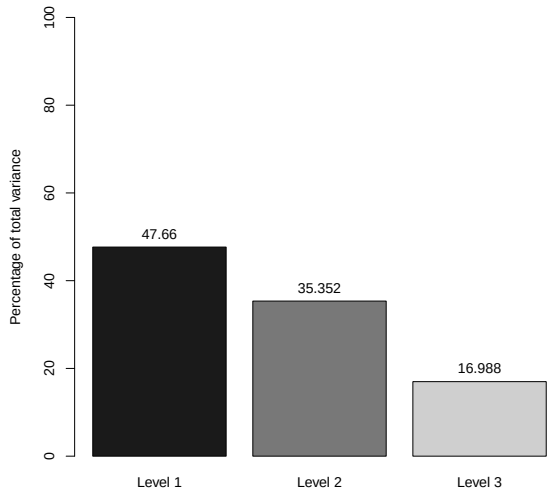


Figure D.1: Variance distribution of Fecundity data

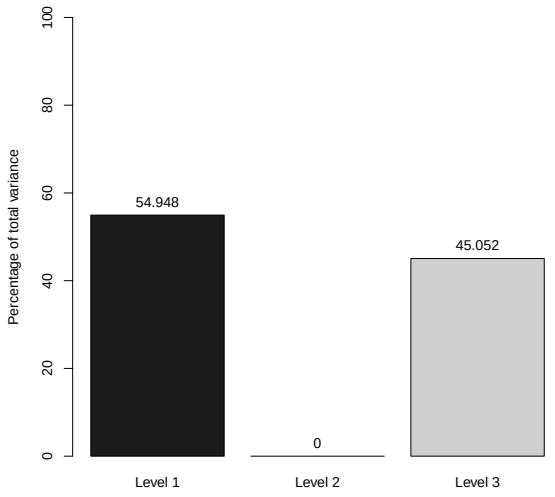


Figure D.2: Variance distribution of Fecundity - fish data

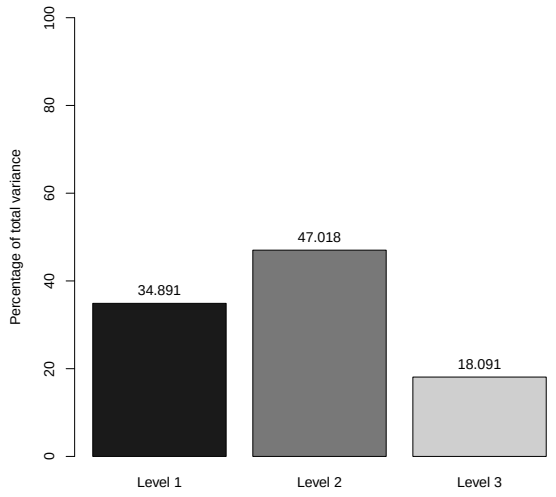


Figure D.3: Variance distribution of Fecundity - crustacea data

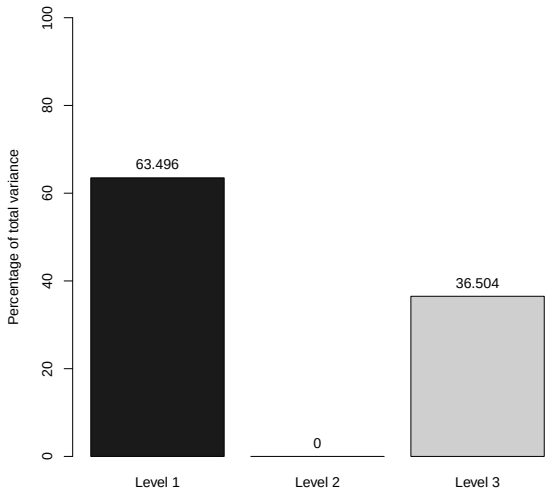


Figure D.4: Variance distribution of Hatchability data

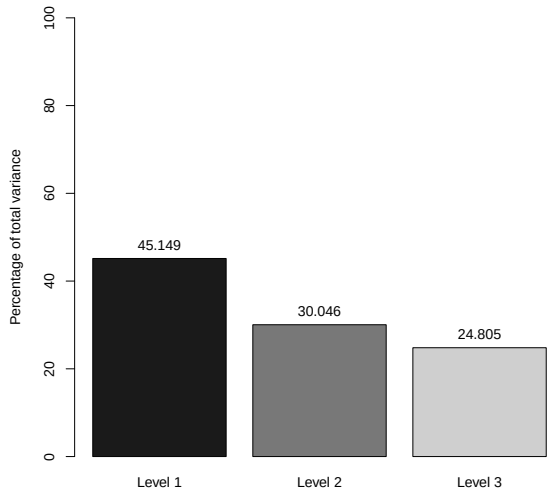


Figure D.5: Variance distribution of Molecular Responses data

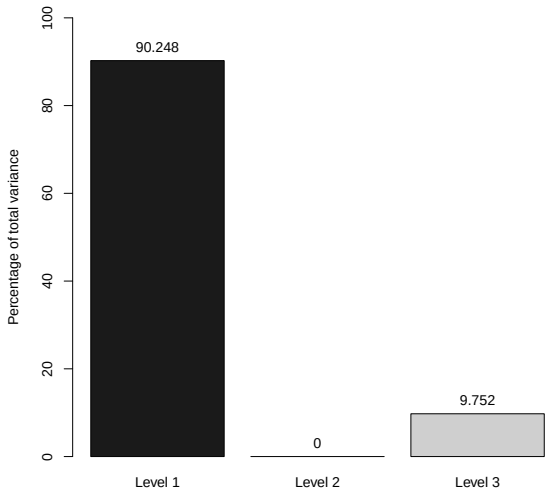


Figure D.6: Variance distribution of Reproductive Behaviour data

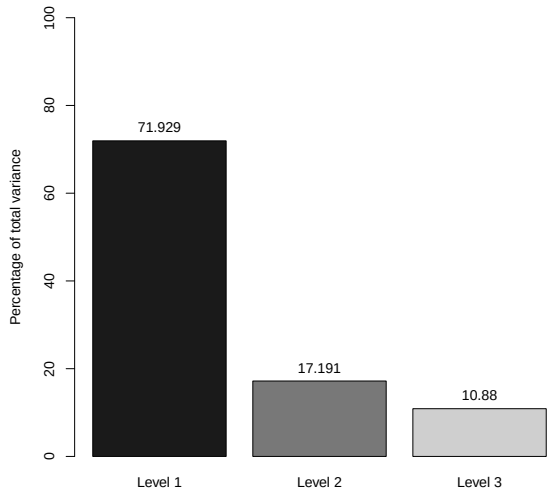


Figure D.7: Variance distribution of Sexual Maturation data

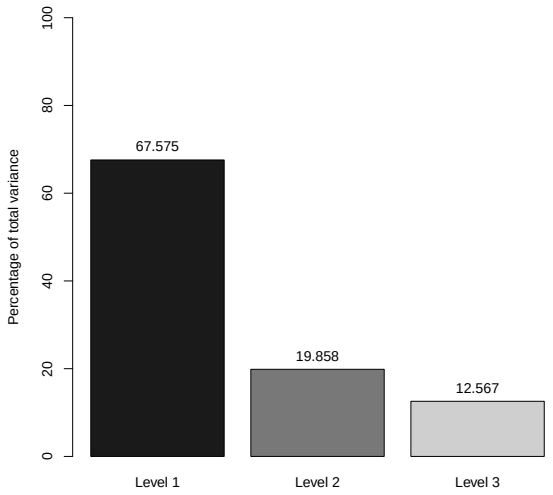


Figure D.8: Variance distribution of Sexual Maturation - fish data

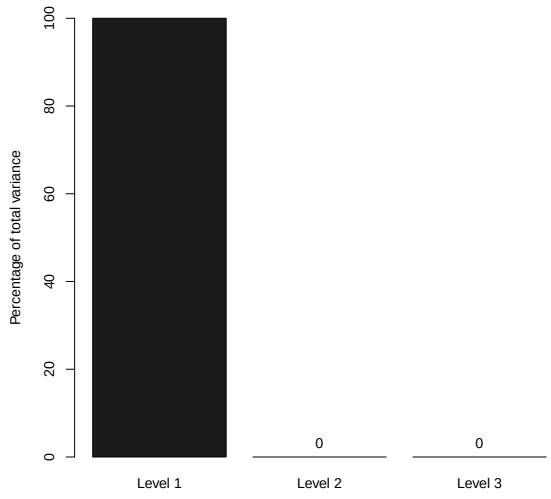


Figure D.9: Variance distribution of Sexual Maturation - crustacea data

Appendix E

Funnel Plots

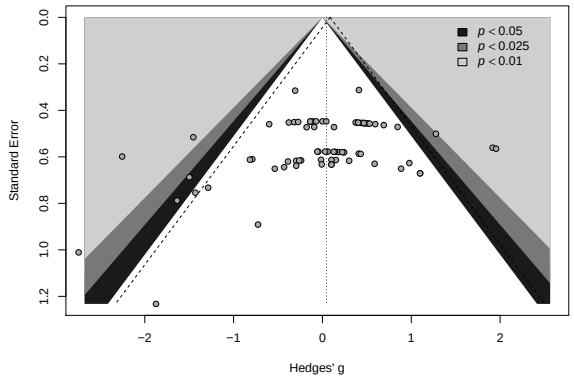


Figure E.1: Fecundity data funnel plot.
Egger's test result: $p = 0.004$*

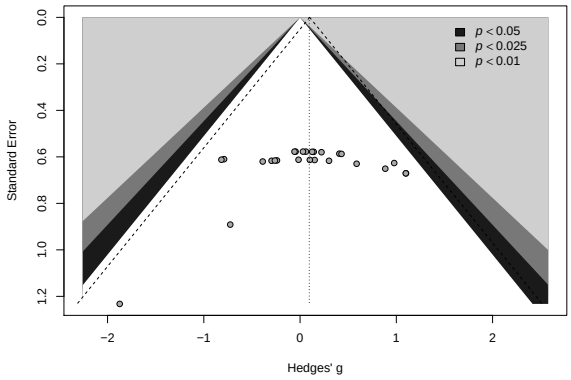


Figure E.2: Fecundity - fish data funnel plot
Egger's test result: $p = 0.1340$

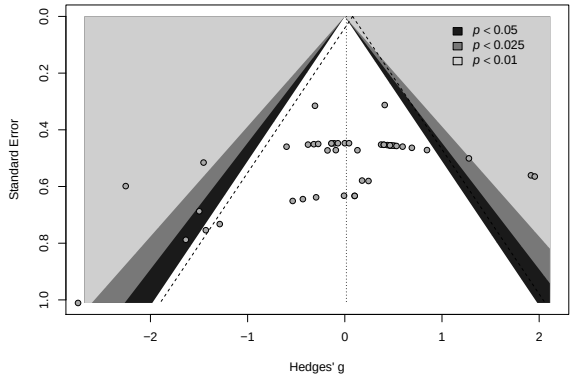


Figure E.3: Fecundity - crustacea data funnel plot
Egger's test result: $p = 0.0017$*

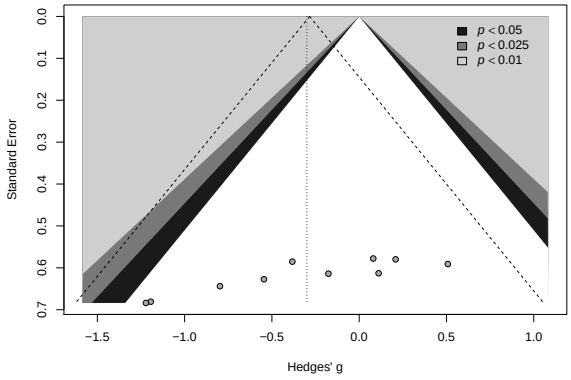


Figure E.4: Hatchability data funnel plot
Egger's test result: $p = 0.0327$*

'*' If the p-value for Egger's test is lower than .05 there is significant funnel plot asymmetry

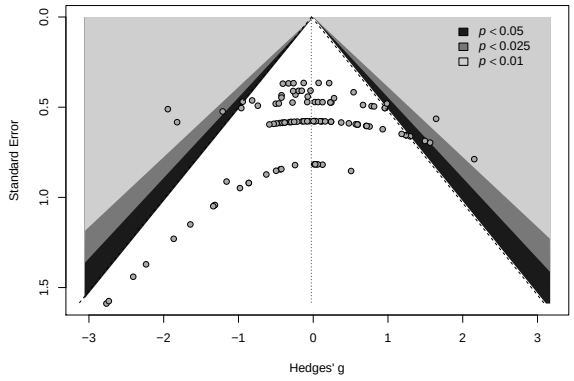


Figure E.5: Molecular Responses data funnel plot
Egger's test result: $p = 0.1196$

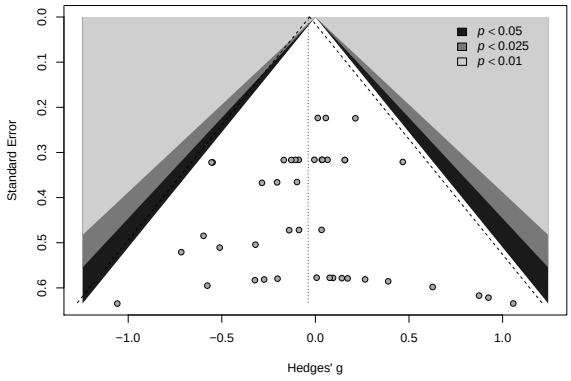


Figure E.6: Reproductive Behaviour data funnel plot
Egger's test result: $p = 0.5566$

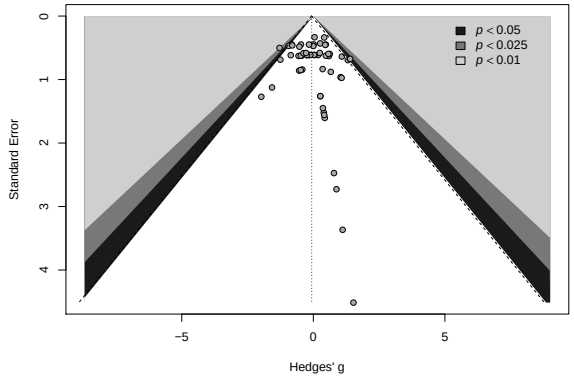


Figure E.7: Sexual Maturation data funnel plot
Egger's test result: $p = 0.8055$

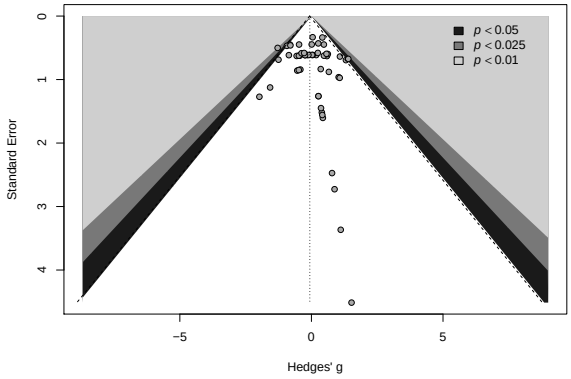


Figure E.8: Sexual Maturation - fish data funnel plot
Egger's test result: $p = 0.9750$

'$*$' If the p-value for Egger's test is lower than .05 there is significant funnel plot asymmetry

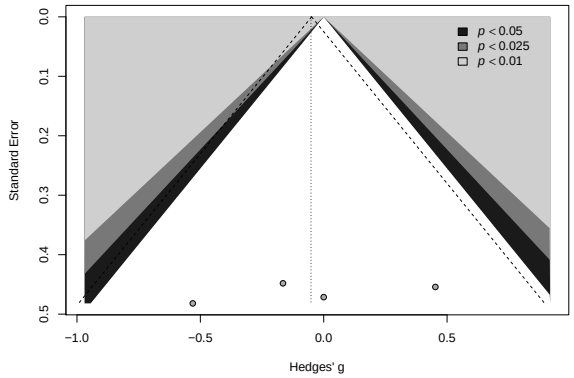


Figure E.9: Sexual Maturation - crustacea data funnel plot

'∗' If the *p*-value for Egger's test is lower than .05 there is significant funnel plot asymmetry